

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:50:18 ; Search time 65 Seconds
(without alignments)
5229.302 Million cell updates/sec

Title: X69111

Perfect score: 7544
Sequence: 1 GARCCTGGGTGCTGCCAGGA.....TTAAACTTTTATAAAAGTT 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3504.5	46.5	3907	5	ABG70822 Mouse myo
2	3504.5	46.5	3907	6	ABG74190 Mouse myo
3	3447.5	45.7	2211	6	ABG72959 Novel hum
4	3322	44.0	7285	6	ABJ38280 pAMG21-RA
5	3215.5	42.6	1679	4	Au07343 1-aminocyc
6	3173.5	42.1	1811	6	Aa37007 Micronono
7	3105	41.2	2601	6	AdA15723 C. elegan
8	3103.5	41.1	1329	5	Au91279 Human NOV
9	3101.5	41.1	1743	6	Abu88255 Novel hum
10	3101.5	41.1	1743	6	Abu90134 Novel hum
11	3101.5	41.1	1743	6	Abu96436 Novel hum
12	3101.5	41.1	1743	6	Abu99045 Novel hum
13	3101.5	41.1	1743	6	Abu98260 Novel hum
14	3101.5	41.1	1743	6	Abu91966 Novel hum
15	3101.5	41.1	1743	6	Abu85270 Novel hum
16	3101.5	41.1	1743	6	Abu00409 Novel hum
17	3101.5	41.1	1743	6	Abu88960 Novel hum
18	3101.5	41.1	1743	6	Abu06456 Novel hum
19	3101.5	41.1	1743	6	Abu95516 Novel hum
20	3101.5	41.1	1743	6	Abu95206 Novel hum
21	3101.5	41.1	1743	6	Abu90754 Novel hum
22	3101.5	41.1	1743	6	Abu93916 Novel hum
23	3101.5	41.1	1743	6	Abu86190 Novel hum
24	3101.5	41.1	1743	6	Abu82845 Novel hum
25	3101.5	41.1	1743	6	Abu07906 Novel hum

26 3101.5 41.1 1743 6 ABU94226 Novel hum
27 3101.5 41.1 1743 6 ABO00099 Novel hum
28 3101.5 41.1 1743 6 ABU87110 Novel hum
29 3101.5 41.1 1743 6 ABU91351 Novel hum
30 3101.5 41.1 1743 6 ABU90444 Novel hum
31 3101.5 41.1 1743 6 ABU97038 Novel hum
32 3101.5 41.1 1743 6 ABO05231 Novel hum
33 3097.5 41.1 2508 6 ADA15721 C. elegan
34 3097.5 41.1 2544 6 ADA15717 C. elegan
35 3082.5 40.9 1300 6 ABU88254 Novel hum
36 3082.5 40.9 1300 6 ABU90133 Novel hum
37 3082.5 40.9 1300 6 ABU96435 Novel hum
38 3082.5 40.9 1300 6 ABU98044 Novel hum
39 3082.5 40.9 1300 6 ABU98259 Novel hum
40 3082.5 40.9 1300 6 ABU91965 Novel hum
41 3082.5 40.9 1300 6 ABU85269 Novel hum
42 3082.5 40.9 1300 6 ABO00408 Novel hum
43 3082.5 40.9 1300 6 ABU88959 Novel hum
44 3082.5 40.9 1300 6 ABO06455 Novel hum
45 3082.5 40.9 1300 6 ABU95515 Novel hum

ALIGNMENTS

RESULT 1
ABG70822
ID ABG70822 standard; protein; 3907 AA.
XX
AC ABG70822;

DT 16-DEC-2002 (first entry)

DE Mouse myocardin associated protein #1.

KW Mouse; myocardin; myocardial infarction; cardiomyocyte;

KW post-mitotic cell; differentiation; therapeutic; gene therapy;

KW heart disease; cardiomyopathy; mortality; heart failure; hypertension.

XX Mus sp.

XX WC020260946-A2.

PD 08-AUG-2002.

PF 21-DEC-2001; 2001WO-US050606.

XX 21-DEC-2000; 2000US-0257716P.

(TEXA) UNIV TEXAS SYSTEM.

Olson EN, Wang D;

WPI; 2002-732693/79.

XX New myocardin polypeptides and polynucleotides, useful for respecifying non-cardiac cells, stimulating cardiac tissue regeneration, and for treating cardiovascular disorders, such as myocardial infarction and hypertension.

Disclosure; Page 146-159; 175pp; English.

XX The invention discloses an isolated polynucleotide encoding a myocardin polypeptide. Myocardial infarction results in the loss of cardiomyocytes, which are post-mitotic cells and generally do not regenerate after birth. Transplanting fetal cardiomyocytes has limitations so identifying new regulators of cardiomyocyte growth and differentiation is an important goal in the search for therapeutics to treat myocardial tissue damage. The polynucleotides, polypeptides and methods of the invention can be used to modulate the phenotype of a non-cardiomyocyte cell to include one or more phenotypic functions of a cardiomyocyte cell, to generate a cardiomyocyte, which comprises introducing into a cardiac fibroblast the myocardin polynucleotide and a promoter which is active in the

fibroblast, stimulate cardiac tissue regeneration which comprises inhibiting the function of myocardin in a post-mitotic cardiomyocyte and for screening for a modulator of myocardin expression. The nucleic acid can also be used in gene therapy to treat a heart disease, including cardiomyopathy, comprising administering a polynucleotide encoding a myocardin peptide or protein or an antisense nucleic acid. The polynucleotide and polypeptide can also be used for decreasing mortality in a subject with heart failure, comprising inhibiting the function of myocardin in post-mitotic cardiomyocytes, increasing the level of myocardin in fibroblasts to generate cardiomyocytes, inhibiting the function of myocardin in post-mitotic cardiomyocytes or increasing the level of myocardin in fibroblasts to generate cardiomyocytes in the subject. The compositions and methods of the present invention are useful for respecifying non-cardiac cells, stimulating cardiac tissue regeneration, for treating cardiovascular disorders, such as myocardial infarction and hypertension, and for screening compounds for various abilities to interact and/or affect myocardin expression or function. The sequence presented is the mouse myocardin associated protein, #1. Note: This sequence is presented as a 3-letter coded protein sequence in the specification, but when changed into the single letter code, forms a DNA sequence identical to that given in AH855224. It, therefore, appears to be the DNA sequence encoding human myocardin 2.

Sequence 3907 AA;

Query Match 46.5%; Score 3504.5; DB 5; Length 3907;
Best Local Similarity 36.8%; Pred. No. 4.9e-148;
Matches 848; Conservative 0; Mismatches 34; Indels 111; Gaps 134;

QY	1	GATCTGGGGTCTCTCC-----AGAAATAAGC-----AAATCTG--GAAGTTAAATGGT	46
DB	754	GGACAGGGGGCACCCCCCATGACTCATCTCTACGCCAAGATCTCTGCAGCAGCAGCACT	813
QY	47	TTT-----GAG-----TGATTTTAA-----AT	64
DB	814	CTTCTCCAGCTGCAGATCCTCAACCAGCAGCAGCAGCACCACAATCACCAGGCCAT	873
QY	65	CTTTGCTGGC-----GGAGAGGCC-----	84
DB	874	CCTGCTGCCCCGCCAAAGTCAAGCAGCAGGAGCCCTGGGAAGCAGCGGACCCCCCAGT	933
QY	85	GGCTCTCC-----CCGCTATCAGC-----GCT	106
DB	934	ACGAGCGCTCTCCACTACCAATAAGCAGCTCCAGCTGGCGGCCCTGGGGCCCTGTGGGCT	993
QY	107	-----TCCTCATTT-----CTTTGAATCGGG-GCTCCG-CGGCTTTC-----	141
DB	994	GGCAGCTCAGAAACAGCACTTCACTGCTGGCAACCGGAGCCCTCCGGCCACCTTGGAA	1053
QY	142	-----GGCTCTAGACC-----AGCCGAG-GAA-----GCTGTGTTT--G	172
DB	1054	CGACATGAAGTGTGCAGAGCTGAAGCAGGAGCTGAAGTTGCGATCACTGCTGCTCGGG	1113
QY	173	CAATTTAAGCGGGCTG--TCAAGGCC-----CAGGGCC-----GGC	206
DB	1114	CACCAAACTGAGCTGATTGAGCGCCTTCGAGCCTATCAAGACCAATCAGCCCTGTGTC	1173
QY	207	GGGG-----CGGGGCCAGCGG--GCC--ATTTTGAATAAG-AGCG--	246
DB	1174	AGGAGCCCCCAAGSCCCTCCGCCACCTCTATCTCTGCACAAGGCTGGCGAGGTGTGTGT	1233
QY	247	TGCTTTCCAGG-----CAGGCTCTATAAGTGACC-GCC-----GGCGCGAGCGT--GGCG-	293
DB	1234	AGCCTTCCAGCGCCCGGCTGAGACACGGGGCCAGCCCTGTGTGGCAGCAGGCGCTGGCTCC	1293
QY	294	-----GCGTTGCAGTCACTGTAG-----CGGA	316
DB	1294	AGCTAGTGTGTGTGGCCACGGTGGCCAGCAGTGGGTGTGTGAAGTTTGGCAGCACGGG	1353
QY	317	CT-----TCT-----TTTGGTTTTCTTCT-----	336
DB	1354	CTCCAGCGCCCCCGTGTCTCCACCCCGCTCGAGCGCTCACTGTCTCAGCAGCGGCGATGA	1413

337	QY	-----CTTTGGGG-----CACTCTGT-----GAC	355
1414	DB	AAACTCCACCCCGGGGACACCTTTTGTGAGATGGTGCATCACTCTGCAGCAGCTGAC	1473
356	QY	-----TCACTCCCCAGCA-----TGAAGCGC-----	376
1474	DB	CCTGAGGCCTCGCCACTGCGATCCTCGTGAAGGAGGAGGCCCCCGGGCGGGTCCGTG	1533
377	QY	-----CTGAGCCC-----GCTGCGCG-----GCTGC-----	397
1534	DB	TTGCTTAGCCCTGTGGGGGGGGCGAGCTAGAGGGCGCGACAAAGNACAGATGCTGCA	1593
398	QY	-----TAGAGGCGGTG-----TGCTGC-----CTGTC	420
1594	DB	GGAGAAAGACAAGACAGATCGAGCGCTGACGCGCATGCTCCGGCAGAGCAGCAGCTGGT	1653
421	QY	GGAAAC-----GAGAGG-----GCATCTG-----GCCAT	438
1654	DB	GGAGCGGCTCAAGCTGAGCTGAGCAGGAAGCGAGCCCGCAGCAGCCGCCCCCGCCCC	1713
439	QY	CGCC-----CGGGGCC-----GAGGAGAGGGCCCGCAGCT--GAGGAGCC--GCTGAGCTT	486
1714	DB	CGCCCCCTTGGCACACCCCGTGAACAGAGAAACAGCTTCTCCAGCTGCCAGCTGAGCCA	1773
487	QY	G-----CTGGACGACATGAACCATGTCTACTCCCGCCT--GCGG-----	523
1774	DB	GCAGCCCCCTGGGCCCCGCTCACCCATTCAACCCAGCCTGCGCGCCCGCCACCAACCA	1833
524	QY	-----GAACTG--GTACCCGAGTCCC-----GAGAGGCA-----	551
1834	DB	CATAGACCTTGTGCTGTGCCCCCGGGGCCCCCGTCCGTGCTGAAGCAGGAAGCCTT	1893
552	QY	-----CTCAG-----CTTAG	561
1894	DB	GCAGCTGACCGAGCGGTGCCCGCCCCCAGTTGTTCTGGGGCTCAGGGCCCCAG	1953
562	QY	CC-----AGGTGG-----AAATCTACAGCGGTTCATCGACTACA-----	596
1954	DB	CCTCATCAAGGGGTTGCACCTCCACC--CTCATCACGCACTCCACAGGACCCACCTTG	2012
597	QY	TTCTC-----GACC-----TGAGGTAGTCTTGGCC-----GAGC-----C	627
2013	DB	TCCTCACCGTGACCAATAGMATGCAGACAGCCCTGSCCTGTCCAGTGGGAGCCCCAGC	2072
628	QY	AGCC-----CTGG-----ACCCCT-----GATGG-----	648
2073	DB	AGCCCTGTCCAGCCTGGCTTCCAGCGCTGCGCCCCCTGCGCCAGATGCACTGGAGC	2132
649	QY	-----CCCCAC-----CTTC-----CCATCCAGAC	669
2133	DB	ACCCACTGCAGCCCTCTTTGGAGCCCCCACTTCTGCTGAAGAGGAACCA--CCTGGC	2191
670	QY	-----AGCC--GAGCTCGCT-----CGGAACTTG--TCATCT--CCA-----	701
2192	DB	TATGAGAAAGCATGAGCCAGCAGCCCAACAGCAGGAATGTTCTTCAAGCCAGCAG	2251
702	QY	-----ACGAC-----AAAGAGCTTTTGCACTGACT-----	729
2252	DB	ATGACAGACCTGTTGACATTCTCATTCAGACGAGAGAAATTTCAGCAGATTTCAAGGAG	2311
730	QY	CGGCGGTGCTCTGACA-----CCT--CCAGAACCGAG-----GTG-----CT	764
2312	DB	CCGSCAT--CCCTGCCAGGAAGAGAACCAATCCCGGAAGACAGTCTGTGGTCCCCCCT	2370
765	QY	GGC-----GCC-----CGTTCTG-----CTGG--GACC-----CGGGAAC	794
2371	DB	GGCAGCAGACCATCACTTCTGTGAGCTCCCCAGGCTGCCCACTCTCTCCAGGCTC	2430
795	QY	CTCTCTGTGC--GGAGGC-----GGAAGGAGGATGG-----GCCCAACT	835
2431	DB	ACCTCCCTCCCTGGACGCTGGAGGACTTCTGTGAGAGCAGCAGGGGGTGTGCCCTGTG	2490
836	QY	-----TCG-----CCCTGCCAC--TTGAC--TTCAACAAATC-----	864

Db 1234 AGCTTCCAGCGGCCGCTGAGACAGCGGCCAGCCCTGGTGGCAGCAGGCGCTGGCTCC 1293
 QY 294 -----GGTTGACAGTCACTGTAG-----CGGA 316
 Db 1294 AGCTGAGTGGTGGTGGCCACCGTGGCCAGCAGTGGGGTGGTGAAGTTTGGCAGCAGCG 1353
 QY 317 CT-----TCT-----TTTGGTTTCTTTCT-----336
 Db 1354 CTCACGCGCCCGTGTCTCCACCCCTCGAGCGCTCACTGCTCAGCAGCGGGATGA 1413
 QY 337 -----CTTTGGGG-----CACCTCTG-----GAC 355
 Db 1414 AAACCTCCACCCCGGGGACACCTTTGGTGAGATGGTGAATCACCTCTGACGAGCTGAC 1473
 QY 356 -----TCACATCCCCAGCA-----TGAAGCG-----376
 Db 1474 CTTGAGGCTCCCACTGAGATCTCTGTGAAGGAGGAGGGCCCCCGGGCTCTG 1533
 QY 377 -----CTGAGCCC-----GGTGGCG-----GCTGC- 397
 Db 1534 TTGCTGAGCCCTGGGGGCGGGAGCTAGAGGGGCGCGACAAGGACAGATGCTGA 1593
 QY 398 -----TACGAGCGGTG-----TGCTGC-----CTGTG 420
 Db 1594 GGAAAGACAAGCAGATCAGGGCGGTGACGCGCATGCTCCGGCAGAAGCAGCTGCT 1653
 QY 421 GGAAC-----GCACTCTG-----GCCAT 438
 Db 1654 GGAGCGCTCAAGTGCAGTGGAGCAGAGAGAGCGAGCCAGCCGCGCCGCCCC 1713
 QY 439 CGCC-----CGGGCC-----GAGGAAAGGCGCGGAGCT-GAGGAGCC-GCTGAGCTT 486
 Db 1714 CGCCCCCTCGGCACCCCGTGAAGCAGGAGAACAGCTTCTCAGCTGCCAGCTGAGCCA 1773
 QY 487 G-----CTGACACATGAACCACTGCTACTCCCGCT-GCGG-----523
 Db 1774 GCAGCCCTGGGCCCGCTACCCATTCAACCCAGCGCTGGCGGCCCGCCAGCCACCA 1833
 QY 524 -----GAACCTG-GTACCCGAGTCCC-----GAGAGCA-----551
 Db 1834 CATGACCCCTGTGCTGTGGCCCCGGGGCCCCCGTCCGTGGTGTGAGCAGGAAGCTT 1893
 QY 552 -----CTCAG-----CTTAG 561
 Db 1894 CGAGCTGAGCCGAGCGGTGCCGCCGCCCGCCAGTTCTTCTGGGCTCAGGGCCCCAG 1953
 QY 562 CC-----AGTGG-----AAATCTACAGCGCTCATGACTACA-----596
 Db 1954 CCTCATCAAGGGGTGACCTCCACC-----CTCATCCGACTCCACAGGACCCACCTTG 2012
 QY 597 TTCTC-----GACC-----TGCAGGTAGTCTCTGCCC-----GAGC-----C 627
 Db 2013 TCCTCAGCGTACCAATAGAAATGCAGACAGCCCTGGCTGTCCAGTGGGAGCCCCAGC 2072
 QY 628 AGCC-----CCTGG-----ACCCCT-----GATGG-----648
 Db 2073 AGCCCTGTCAGCTGGCTCTCCAGCGCTGCCCTCTGTCCAGATGGAAGCTGAGC 2132
 QY 649 -----CCCCCAG-----CTTC-----CCATCCAGAC 669
 Db 2133 ACCCACTCAGCCCTCTTTGGACCCCACTTCTCTGCTGAAGAGGAACCA-CTTGGC 2191
 QY 670 -----AGCC--GAGTGGCT-----CCGGAACCTG--TCATCT--CCA-----701
 Db 2192 TATGAGGAAGCATGAGCCAGCAGGCCCAACAGCAGGAAATGTTCTCAAGCCAGCAG 2251
 QY 702 -----ACGAC-----AAAAGAGCTTTTGCACTGACT-----729
 Db 2252 ATGAGCAGCTGTTTGAATTCATTCTATTGAGCGGAGAAATTTGAGAGATTTCAAGGAG 2311
 QY 730 CGGCGGTGTCTTGACA-----CCT--CCAGAACGAG-----GTG-----CT 764
 Db 2312 CCGCCAT-CCCTGCCAGGAAGAGAGCCATCCCGGAAGACAGTCTGTGGTCCCCCT 2370

QY 765 GGC-----GCC-----CGTTCTG-----CCTGG--GACC-----CCGGGAAC 794
 Db 2371 GGCAGCAGCAGCATCACTTCTGTGAGCTCCCGCAGGCTGCCCCACCTCTCCAGGCTC 2430
 QY 795 -CTCTCTGCG--GGAGCC--GGACGGCAGGGATGG-----GCCCAACT 835
 Db 2431 ACCCTCCTCCCTGAGCGCTGAGGACTTCTTGGAGAGCAGCAGGGGGTGCCTCTGCT 2490
 QY 836 -----TCG-----CCCTGCCAC--TTGAC--TTACCAAAATC-----864
 Db 2491 GACCAGTGGCATGACGGCCAGAGCCCTTCCCTCATTGACGACCTCCATAGCCAGAT 2550
 QY 865 -----CCTTCTGGAGACTAACTGTGTGCTCAGGAGC-----GAA--GG 902
 Db 2551 GCTGAGCAGCACTCCCATCTCTGA--CCACCCCGCTCACCCTAGGACACCTCGGAATTGC 2609
 QY 903 ACTGTGAACCTGTAGCTGAAGAGCCA-----GAGCTAGCT-CTGGCCACC-----947
 Db 2610 ACTTTGTCTTGAACCCAGCAGCACCATGGCCCTGGACCTGGCTGATGGCCACCTGGACA 2669
 QY 948 -----AGCTGG-GC-GACGTCA-----CCC-TGCT-----CCACC-----975
 Db 2670 GCATGGAATGGCTGAGCTGTCTGAGTGGTCCGCTGAGCCTAGCCCTCCCTCAGCA 2729
 QY 976 CCAC--CCCCAAG--TTCT--AAAGTCTTTTTCAGA--GCGTGGAGGTGTGGAAGGAGT 1025
 Db 2730 CCAAGCCCCAGCCTCTTCTCCACAGACTTCTCGATGGCCATGATTTCAGCTGCACT 2789
 QY 1026 GG-----CTGCT-----CTC-CAAACT-ATGCCAAGCGGGCGGACAGCTGG-----1065
 Db 2790 GGGATTCTGCTGTAGCTCTCTGGCTCAAGACGGGTGGGAAGGGCTGGGAGCCAGG 2849
 QY 1066 --TCTTC-----TGG-TCTCTTTGGAGA-----AAGGTCTCTG-----TTG- 1097
 Db 2850 GTACTCCAAATCGTGGCTCTCTCGGTGATTCGGCCCTCCACATGGTTGTGACTTTGA 2909
 QY 1098 -----CCCTGATTTAT-----GAACTCTATAATAGATA-----TAT 1129
 Db 2910 CAATCACAGCCCTGCTTTTCCCTTCCCTGGAGGCTAGAACAGAGAAGCCCTTACTCC 2969
 QY 1130 AGGTTTGTACTTTTTCAGG--GAGTGACTTTCTGTAACAATGCG--ATGTATA 1183
 Db 2970 TGGTTCACTGGC-----ACGACGGCAGAGGAGAGCAGCTGTCAAGAGCAGCCCTGGCTC 3025
 QY 1184 TTAAGT-----TTTT--ATAAAGTT 1203
 Db 3026 TCACGCTGGGTTTGGACACACGCT 3051

RESULT 3

ABG72959

ID ABG72959 standard; protein; 2211 AA.

XX

AC ABG72959;

XX

DT 04-APR-2003 (first entry)

XX

DE Novel human phospholipase associated protein.

XX

KW Human; phospholipase; phospholipase C subfamily; gene therapy;

KW pharmacogenomic analysis; transgenic; ribozyme design;

KW antisense technology.

XX

OS Homo sapiens.

XX

PN US2002155572-A1.

XX

PD 24-OCT-2002.

XX

PF 14-MAR-2002; 2002US-00096961.

XX

PR 14-SEP-2000; 2000US-0232632P.

18-DEC-2000; 2000US-00738884.

(PEKE) PE CORP NY.

Guegler K, Beasley EM, Ketchum KA, Di Francesco V;

WPI: 2003-182638/18.

Novel isolated human phospholipase peptide and gene encoding the protein, useful for diagnosing, treating and preventing disorders associated with aberrant or unwanted phospholipase expression.

Disclosure: page 20-25: 57pp: English.

The invention describes an isolated human phospholipase polypeptide (I), its allelic variant or orthologue. (I) and the polynucleotide encoding it (II) are useful as models for the development of human therapeutic targets, for identifying therapeutic proteins, and as targets for development of human therapeutic agents that modulate phospholipase protein activity in cells and tissues. (I) is useful in assays to determine the biological activity of the protein and levels of the protein in biological fluids, as markers for tissues in which the corresponding protein is preferentially expressed, to identify binding partners to develop a system to identify inhibitors of binding interaction, and in the treatment of disorders characterised by an absence of or unwanted expression of the protein. (I) is also useful in drug screening assays, to identify compounds that modulate phospholipase protein activity of the protein. (I) is also useful for screening a compound for its ability to stimulate or inhibit interaction between the phospholipase protein and a molecule that normally interacts with the phospholipase protein, in pharmacogenomic analysis, and for treating a disorder characterised by an absence of, in appropriate or unwanted expression of the human phospholipase protein. (II) is useful for constructing recombinant vectors, host cells and transgenic animals, for expressing antigenic portions of the human phospholipase proteins, for designing ribozymes, for monitoring the effectiveness of modulating compounds on the expression or activity of a human phospholipase protein gene, as an antisense construct to control human phospholipase protein gene expression in cells, tissues and organisms, and for gene therapy. This is the amino acid sequence of a novel human phospholipase C. Subfamily phospholipase associated protein. Note: This sequence appears in the sequence listing as SEQ ID number 1 but is not described in the specification

Sequence 2211 AA;

Query Match 45.7%; Score 3447.5; DB 6; Length 2211;
Best Local Similarity 36.9%; Pred. No. 1.1e-145;
Matches 809; Conservative 0; Mismatches 355; Indels 1027; Gaps 124;

4	QY	CTGGGGTCTCTCCGAGGAAAAAGCAAAATCTGGAGTTAA--TGTTTT--GAGTGATTTT	59
37	Db	CTGGGGGGCTCCC--GGCTCCGAAAGATCGCTCGCGCAGCTGGCAAGAGAGCGGTGT	94
60	QY	TAAATCTCTTGTGGCGAGAGGGC-----CGCTCTCTCCCG-----	95
95	Db	ACCGGC--TGCAGAGGA-CGGCTGAGCGTGTGGTTCCAGCGCGCATCCCGCGTGGC	151
96	QY	GTATCAGCGTTCCTCAT-----TCTTTGA-----ATCGC-----GGC-----TCGG	133
152	Db	CATCGCAGCATATTTCTTCGTGCAGACATGAGCGGTCCGCGAGGCGCCACGAGTCGG	211
134	QY	C-GGTCATTCCGGC--TCAG-----ACC-----AGCCGGAG-GAAGCCT--GTTTGCAAATTT	178
212	Db	AGGGCTTCGGCGCTTCGGGGTGCTTCGGCGCAGCGCGTGCCTCACCATCGCTTCA	271
179	QY	AAG-----CGGG--CTGTGAACCCCGAGGC-----CGGC---GGG	209
272	Db	AGGGCGCGCGAAGACCTGGACTCGGGGCGCCACGCGCTGAGAGCGCAGCGCTGGG	331
210	QY	GGCGGGG-----CCGAG-GCG--GGCCATTTTGAATA-----AGAGGGGTG	248
332	Db	TGCGGCTGTGACCAAGCTCCGCGCGCTGGAGCGCATGACGAGCGCGAGCGGTAG	391

249	QY	-----CCT-----TCCAGCAGGCTCTATTAAGTGAACCGC-----	277
392	Db	ACCACCTGGATCCCACTCCTATCTGCACCGGGCTGATCCAAACGAGACACAAGATGAGCT	451
278	QY	-CGCGCGAGCGTGC CGCGGTTG--CAGGTCACCTGTAGC--GGACTCTTTTGTG-TGTTT	329
452	Db	TCAAGAGAGATCAAGAGCGCTGCTGAGAAATGGTCAACGTGGACATGAACGACATGTACGCT	511
330	QY	TCCTTCTCTTTGGGCG-----ACCTCTGGACTCACTCC--CCAGCATGAAGCGCTGAGC	382
512	Db	ACCTCTCTTTCAAGAGAGTGTGACCACT--CCAAACAACGACCTCTAGAGGGGCTGAGA	568
393	QY	CCGG-----TGC-GGGGCTGCT-----	398
569	Db	TCGAGGAGTTCTCTGCGCGGGCTGCTGAAGCGCCGGAGCTGGAGGAGATCTTCCATCAGT	628
399	QY	-----ACGAGG-CGGGTGTGCTG-----CCTGTCCGAAACG--	426
629	Db	ACTCGGCGAGGACCGCGTGTGCTGAGTGCCTCTGASCTGCTGGAGTTCTCTGGAGGACCGG	688
427	QY	-----CAGTCTGGCATCGCCCG-----GGCC--GAG-----G	453
689	Db	GCGAGGAGGGCGCCACACTGGCCCGCGCCGACGACTCATTCAGACCTATGAGTCAACG	748
454	QY	GAAGGGCCGCGAGC-TGA--GGAGCGGCTG--AGCT-----	485
749	Db	AGACAGCCAAAGCAGCATGAGCTGATGACACTGGATGGCTTCATGATGTACCTGTGTGCG	808
486	QY	-----TGC--TGGAC-----GACATGAAC--	503
809	Db	CGGAGGGGACTGCTTTGGACAACACCCACAGTGTGTCCAGGACATGAACACGACCCC	868
504	QY	-----ACTGCTACTC--CGCGCTGGGA-----ACTGCTACCC--	538
869	Db	TTGCCCACTACTTCACTCTTCTCTCCCAACACCTATCTGACTGACTCCAGATTCGGGG	928
539	QY	GTCCCGGAGGCACTCAG-----CTTTAG--CCAG-----GTGGAA	571
929	Db	GGCCG-AGCAGCACCGAGGCGCTATGTTAGGGCGTTTGCCAGGATGCCGCTGGTGAG	987
572	QY	AT--CCTAC-----AGCGCGTCATCGA-----C	592
988	Db	CTGGACTGTCTGGAGGGCCAGAGGGGAGCCCGTCACTATCATGSCCATACCTCACC	1047
593	QY	TAC--ATTCTGACCTGCAGGTAGTCTCTGGCCGAGC-----CAG-CC	631
1048	Db	TCCAAGATTCTTCTCGGGAGCTGGCCCAAGCGGTGCGCAGCATGCTTCAAGCTGTCC	1107
632	QY	CCTGGACC-----CCCTG-----ATG	647
1108	Db	CCATTACCTGTCTCATCTATCTCCCTGGAGAACCACCTGCGGGCTGGAGCAGGCTGCCATG	1167
648	QY	GGCC-CGACCT--TCCCATCCAGACAGCGGAGCT--CGCT-----	692
1168	Db	GCCCGCCACTCTGCACTCATCTCTGGGAGACATGCTGTGTGACACAGGCGCTGACTCCCCA	1227
683	QY	----CC--GGAACTTGTCATCTCC--AAGCAAAAGGAGC--TTTGTG-----	720
1228	Db	AATCCCGAGAGC-TGCCATCCCCAGAGCAGCTGAAGCGCGGGTCTCTGTGTGAAGGGAAA	1286
721	QY	-----CCACTGACTC-----GGCGGTGCTCTGAC-----	744
1287	Db	GAAGTCCCGCTG-CTCGGAGGAGGATGCGCGGCTCTGTGATCGGAGGAGGGGG	1345
745	QY	-----AC-----CTCCAGAGC-----	756
1346	Db	AGGAGGATGACGAGGAGGAAGAGGAGTGGAGGCTGCAAGCGCAGAGCGGCTGGCCA	1405
757	QY	-----CAGGTGCTGGC-GCCC-----GTTCTGCTGGGACCCCG-----	792
1406	Db	AGCAGATCTCCCGAGGCTGTGGCCCTCTGGTGTGTACTGTCACGCAACCGCTCTGGGA	1465

Db 3040 GGCAATCGCCATGAGTACAGACGAGATCTCGCGTGGGCAATGCGCGCTTGGAGCTGGC 3099
QY 388 GCG-----CGGCTGTACAGAGGGTG-TGCT-----GCCTGT-----CGAAGC--CAG 429
Db 3100 GAACATTCGGCTGGCGGAGCCCTGATGCTCTTCGTCAGATCATCTGATCGACAAG 3159
QY 430 TCTGGGCATCGCCCGCGCGCGAG-----GGAAGG 458
Db 3160 ACCGGCTTCCATCCGAGTAGCTGCTCGCTCGATGCGATGTTTCGCTTGGTGTGGAATGG 3219
QY 459 GC-----CCGG--CAGC--TGAGGAGCGCTG-----AG-CTTGTGGA----- 492
Db 3220 GCAGGTAGCGGATCAAGCGTATGCGAGCGCGCGCATTTGATCAGCATGATGATATCTT 3279
QY 493 --CGACATGAAC--ACTG-----CTACTCCGC-CTGCGGGAATGTGTAC 533
Db 3280 CTCGGCAGGAGCAAGGTGAGTACAGAGGATCTTCCCGGCACTTCCGCCAATAGCAG 3339
QY 534 CCGG---AGTCCG--AGAGCA-----CTCAGCTTACCGAG-----GT 567
Db 3340 CCAGTCCCTTCCCGCTTCACTGACAACTGCGAGCAGCTGCGCAAGGAAAGCGCGCTGT 3399
QY 568 GGAATCC-----TACAGCGCGTCAATCGAC-----TACATCTCGAC-CTG--CAGGTA 613
Db 3400 GGCAGCAGATAGCGCGCTGCTGCTGCTGCAATTCATTTCAGGACACCGGACAGTTC 3459
QY 614 GTCTGGCC-----GAGCC---AGCCCT-----CGACCC----- 641
Db 3460 GGTCTTGACAAAAGAACCGCGCGCCCTGCGCTGACAGCGGAAACAGCGCGGCAATCAGA 3519
QY 642 -----CTGATG-GCCC-----ACCTTCCCATCCAGACAGCGC- 674
Db 3520 GCAGCGATGTCTGTGTGTCGCGCAGTCATAGCGGAATAGCTCTCCACCAAGCGCGCG 3579
QY 675 -----AGCT-----CGCTCGGAACTGTG----- 694
Db 3580 AGAAGCTGCTGCAATCCATCTGTGTCATATGCGAAGCAATCTCATCTGCTCTTGT 3639
QY 695 -----ATCTCAACGACAAAGAGCTT-----TTGCC-----AC 724
Db 3640 ATCTGATCTTGATCCCTGCGCCATCAGATCTTGGCGGCAAGAAAGCATCCAGTTAC 3699
QY 725 T-----GACT---CGGCGGTGTC-----CTGACACTCCAGAACGAGGT 761
Db 3700 TTTCAGGGCTTCCCAACCTTACAGAGGCGCCCGAGCTGGCAATTCGGTTCGCT--T 3757
QY 762 GCTGG-----CGCCCGTCTT-----GCC---TGGGACC-----CCGG 791
Db 3758 GCTGTCCATAAAACCGCCAGTCTAGCTATGCGCATGTAAGCCCACTGCAAGCTACCTGC 3817
QY 792 AACCTC-----TCCTG-----CCGAAGC---CGGACGCGAGG 821
Db 3818 TTTCCTTTGGCTTGGTTTTCCCTTGTCCAGATAGCCAGTAGCTGACATTCATCCGG 3877
QY 822 GATGGGCCCAACTTCGCGCTGCCACTTGACTTC-ACCAATTC-CTTCTC---GGAGA 876
Db 3878 GGTTCAGCAGCGTTT-----CTGGGACTGGCTTCTACGTGTTCGCTTCTCTTTCAGCAGC 3932
QY 877 CT---AAACCTG-CTGCT--CAGGAGGAGGAC-----TGNGAACTTGTAGC--- 918
Db 3933 CCTTGGCCCTTGAAGTCTTGGCAGCGTGAAGCTACATATATGTATGATCGG--GCAAT 3990
QY 919 --CTGAAGACCAAGAGTACTCTGCGCAACAG-----CTGGG-CGACGTC----- 961
Db 3991 CGCTGAATATTC--TTTGTGCTCCGACCATCAGGACCTGAGTCGCTGCTTTTTCGTGA 4049
QY 962 -----ACCTGTCTCCAC-----CTGGG-CGACGTC----- 974
Db 4050 CATTCAGTTCGCTGCGCTACGCTCTGCGAGTGAATGGGTAATGGCACTACAGCG 4109
QY 975 CC-----CACCCCAAGTTCT-----AAGGTTCT 998

Db 4110 CCTTTTATGATTCATGCAAGAAACTACCATAATACAGAAAAAGCCGTCACGGCTT 4169
QY 999 TTCAGAGCG-----TGGAG-----TGTTG-----RAGGA 1023
Db 4170 CTCAGGCGCTTTTATGCGCGGTCTGCTATGTGGTGTATCTGACTTTTGTGTTTCAGCA 4229
QY 1024 GTGGCTGCTCTC-----CA 1037
Db 4230 GTTCTGCTCTCTGATTTTCCAGTCTGACCACTTCGGATTATCCGTCACAGGTCAITCA 4289
QY 1038 AACT-----ATGCC-----RAGGCGCG-----CAGAGCTG-----GTCTT--- 1069
Db 4290 GACTGGTATGACCCAGTAGGAGCGGTATCATCMACAG-GCTTACCGCTTCTACTG 4348
QY 1070 -----CTGCTCTCC-----TTGG----- 1082
Db 4349 TCGAAGACGTGCGTAACGATATGCTATGCTCTCCCATCGAGATAGGAACTGCCAGGCA 4408
QY 1083 -----AGAAAGGTTCTGTTGC-----CCT---GATTTA----- 1107
Db 4409 TCAATATAACGAAGGCTCAGTGAAGACTGGGCTTTCGTTTATCTGTGTTTGTGTC 4468
QY 1108 --TGAACCTCTAT---AATAGAGTATAT-----AGTTTGTACCTTTTAA--- 1148
Db 4469 GGTGAACGCTCTCTCTGAGTAGGACAAATCCCGCGGAGCGGATTTGAACGTTGGAAGCA 4528
QY 1149 -----CAGGAAGGTGACTTCTGTAT-----ACATGCGATGTAT---ATTAAAC- 1189
Db 4529 ACGGCCGAGGCTGGCGGAGGACGCCGCCCATAACTCCAGGATCAATTAAGCA 4588
QY 1190 -----TT-----TTTATAAAGTT 1203
Db 4589 GAAGGCCATCTTCAGCGATGGCTTTTTCGCTTTTACAAACT 4631

RESULT 5

AAU07343
ID AAU07343 standard; protein; 1679 AA.
XX
AC AAU07343;
XX
DT 04-DEC-2001 (first entry)
XX
DE 1-aminocyclopropane carboxylate (ACPC) synthase #12.
XX
KW 1-aminocyclopropane carboxylate synthase; ACPC synthase; brain trauma;
KW excitatory neurotransmission; neurodegeneration; stroke; nerve damage;
KW neurodegenerative disease; Alzheimer's disease; depression; epilepsy;
KW alcohol abuse; cognitive function; memory; learning impairment; human.
OS Homo sapiens.
XX
PN WO200168879-A2.
XX
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-EF002857.
XX
PR 14-MAR-2000; 2000US-0189086P.
PR 05-APR-2000; 2000US-0194702P.
XX
PA (PARB) BAYER AG.
XX
PI Ramakrishnan S;
XX
DR WPI; 2001-550286/61.
XX
PT Isolated polynucleotide encoding a human 1-aminocyclopropane-carboxylate
PT (ACPC) synthase, useful for treating brain trauma and neurodegenerative
PT disease (e.g. Alzheimer's disease, depression, epilepsy).
XX
PS Claim 1; Page 231-237; 242pp; English.
XX

The invention relates to reagents and methods for regulating excitatory neurotransmission, and to prevent neurodegeneration. The method involves the use of an expression vector or a reagent that modulates the activity of a 1-aminocyclopropane-carboxylate (ACPC) synthase polypeptide. The reagent is useful for modulating the activity of an ACPC synthase in a disease such as stroke, a nerve damage or a neurodegenerative disease. The ACPC synthase polypeptide, polynucleotides and modulators are also useful for treating brain trauma and neurodegenerative disease (e.g. Alzheimer's disease, depression, epilepsy). The ACPC synthase modulators are also useful for treating alcohol abuse and improve cognitive function and memory of patients with learning impairment. The present sequence represents the amino acid sequence of human 1-aminocyclopropane-carboxylate (ACPC) synthase #12, used in the method of the invention

Sequence 1679 AA;
XX
SQ

RESULT 6	AAE37007	AAE3	23-0	07-2
ID	AAE3			
XX				
AC				
XX				
DT				
DT				
XX				

DE	Micromonospora carbonacea polyketide synthase (PKS) type I #13.
KW	Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme.
XX	Micromonospora carbonacea.
XX	
XX	Key Location/Qualifiers
FH	Misc-difference 452..1811
FT	/note= "Encoded by GCC"
XX	
PN	CA2391131-A1.
PD	19-NOV-2002.
XX	
PF	26-JUL-2002; 2002CA-02391131.
XX	
PP	26-JUL-2001; 2001US-0307629P.
PR	(ECOP-) ECOPIA BIOSCIENCES INC.
PA	
PB	Yang X, Staffa A, Farnet CM;
PI	WPI; 2003-3493556/33.
PS	N-PSDB; AAD5823.
DR	
DT	
XX	
PT	Novel isolated polypeptide involved in biosynthesis of macrolides by microorganisms, useful for biosynthesis of macrolides by microorganisms, preferably for biosynthesis of rosaramicin.
PT	
PT	Claim 13; Page 183-189; 206pp; English.
XX	
CC	The invention relates to genes and proteins involved in the biosynthesis of macrolides by microorganisms. In particular it relates to the nucleic acids forming the biosynthetic locus for rosaramicin (a 16-member macrolide antibiotic) from Micromonospora carbonacea. The invention is useful for the biosynthesis of macrolides by microorganisms. It allows direct manipulation of macrolides and related chemical structures by chemical engineering of the proteins involved in the biosynthesis of rosaramicin. It is useful to catalyse certain biochemical reactions, in vitro or in vivo, to direct or enhance the synthesis or modification of a polyketide, polyketide substrate or its precursor. The present sequence is M. carbonacea polyketide synthase (PKS) type I protein. (Updated on 23 -OCT-2003 to standardise OS field)
CC	
XX	
SQ	Sequence 1811 AA;
	Query Match 42.1%; Score 3173.5; DB 6; Length 1811;
	Best Local Similarity 46.9%; Pred. No. 1.4e-133;
	Matches 683; Conservative 0; Mismatches 385; Indels 387; Gaps 99
QY	2 ATCTGGGCGTGTGCCAGGAAAAGCAATTCTTGAACTTAATCGTTTGAGTGATTTTATA 61
Db	453 ATGCACAGACACCCTCCGAGCTGAACGGATACTCGACG-----C 490
QY	62 AATCTTGTGTGGCGGAGAGGCC---CGC-----CT--CTCCCGGTATTCAGCGTTCTCT 110
Db	491 GATCTCTTCG---CGAGGAGACGAGCGCGGGAGCTGCGCGGCTTGCCTCTCTCTCTCT 546
QY	111 CATTTCTTTGAATCCG---CG--GCTC-----CG-----TCTTCGGCGTGCAGAC 151
Db	547 CTATCCCGG---CCGTGACGGTGCACAGGACGAGACGGGATGTTCTCTGGGC--CTTCC 601
QY	152 CAGCCGGAGGAG--CCTGTTTGCATTTAAGCGGCTGTGAAGCCGACG---CCCG- 204
Db	602 CGCCACGAGGAGAGCACCG--CGCAA-----GTCTGTGCACAGGAGAGTGCCTGG 651
QY	205 -GCGGCGGCGCGGCGCGAGCG-CGCCCATTTTGAATAAAGAGCGGT-GCCTTCAGGACG 261
Db	652 TGCCCGAGCTGGGCCCGCGGAGGCCCT-----CCTCGCGTCTCTGCCCAG 697
QY	262 CTCCTATAAG-----TGACCGCCGCG--CGAGC--GTGCGCGCGTTTCAGGT 304
Db	698 CTCGGTCACTACAACACGGTCTGTCGTCTGTTGTGAGCCGCTGCCACCTT-CGGCT 755

Db 1676 GAGTACAAATGTTGACGTCGAAGGAGCGAATTTGTCGACAGCGGTCAAGCGCGACGCG 1735
Qy 760 -----GTGCTGGG-----CCC---GTCTGCTG-----GACCC-----787
Db 1736 AACATCGAATGTCGAGTTATTAATTCATAAAATCTGCAAAACGACACTTGCACG 1795
Qy 788 -----CGGGAAC---CTCTC---CTGC-----CGGAAG-----809
Db 1796 ACTTCATCGGGACCTACTCTGCTGACGGTTAGTTTCATATTTGATCGGACAGCGGC 1855
Qy 810 -----CGGACGG-----CAGGGAT-----824
Db 1856 TTCTACTTTCTCAATATATTTTCTCGCCAGCTCGTCGTAGTTTATCATGATCTCA 1915
Qy 825 -----GGGCC---CAACTTCG-----CCC-----TGCCC 846
Db 1916 TTCTGGATCAATCGTACTCGCGCTTCGCGAACCTAATCGGTACGATGACGCTGCTC 1975
Qy 847 ACT-----TGAC-----TTACCA-----860
Db 1976 ACTGAGACTCATCTTATGACCGGAACCAATGACAGCTTCCACCACTGTCCTATGTA 2035
Qy 861 -----AATCC-----CTTC-----CTGG-----AGACTAA 880
Db 2036 GCCGTTGATGATTTCTCGGTTTCTGCTATCTTCTGTTTATCTGCGCTTGCATCGATC 2095
Qy 881 ACCTGG-----TGCTCAGAGCGAAGGACTGTGAACIT-----GTAGCCTCAAG-AGCCA 929
Db 2096 GCCTGTGCTCTACTCAAAAGAGAGACGAGGATCGTCGAGAGAGAGAGAGAGAGC 2155
Qy 930 GAGCT-----AGCTCTG---GCCACAG-----CTGGGGAGCTCACCT-----966
Db 2156 GAGCATAACTGCTCGCGGACACCTGATATTTCTCAGAGCTCGCGCTTGGCGAATGC 2215
Qy 967 -----GCTCCACC-----CCACCCCAAGTTCTTAAGTGCTTTTTCAGA 1004
Db 2216 ACATGCAACGCGCTCCACCTCGATCATCGCGCTCATCAGAGCTGCAATCGATCTGT 2275
Qy 1005 G-----CGTGGAGTGTGGAAGAGTGC-----TGCTC-----1033
Db 2276 GTCACTCAGTCAATGACAT-CTCAGCGTGGCGGTTTCTCTGTGTTTCATCTT 2334
Qy 1034 -TCCAACTATCCCAAGCGG---CGGCAGAGCTGCTCTCTGTGCTC-CC-----TTGG 1082
Db 2335 GTTCAACACCTCTCTGCTGATTTCTACTGTACAAATCCAAAGCTCTGCGGTATATAG 2394
Qy 1083 AGAA-----AGGTT---CTGTTC-----CCTGA---TTTATGAA-CTCTAT-----1117
Db 2395 TGAACAGAGGTGACCGTGTGATGCTCCAGACTTCATTAATCTCAATCCAACTTCCT 2454
Qy 1118 -----AATGAGTATAGGTT---TTGTAC-----CTTTTTC-AGGA 1153
Db 2455 CATCATTTCCATTTCCGATATCTCTTTTCTTTCACAGAGCCTTTTTCGTTTTTTT 2514
Qy 1154 AGGTGACTTTCTGTAAC-----AATGC---GATG-----TATATTAAC 1189
Db 2515 TATGATTTATTTTACGATTTTATGATTAATGACAGATGCTCATGCTCAATTAAT 2574
Qy 1190 TTTTATAA 1198
Db 2575 TTATTTAA 2583
RESULT 8
AAU91279
ID AAU91279 standard; protein; 1329 AA.
XX AC AAU91279;
XX AC AAU91279;
XX DT 18-JUN-2002 (first entry)
XX XX
DE Human NOV3a protein.

XX Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis; diabetes;
KW cell signal processing; metabolic pathway modulation; inflammation;
KW autoimmune disorder; scleroderma; transplantation; allergy;
KW systemic lupus erythematosus; haemophilia; Alzheimer's disease;
KW graft versus host disease; Lesch-Nyhan syndrome; periodontitis;
KW pancreatitis; musculoskeletal disorder; Parkinson's disease;
KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;
KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;
KW growth disorder; reproductive disorder; lung disease.
XX Homo sapiens.
XX OS
XX WO200216600-A2.
XX PN
XX 28-FEB-2002.
XX PD
XX 27-AUG-2001; 2001WO-US026518.
XX PF
XX 25-AUG-2000; 2000US-0227800P.
XX PR 25-AUG-2000; 2000US-0228205P.
XX PR 25-AUG-2000; 2000US-0228324P.
XX PR 30-AUG-2000; 2000US-0228997P.
XX PR 30-AUG-2000; 2000US-0229185P.
XX PR 01-SEP-2000; 2000US-0229780P.
XX PR 01-SEP-2000; 2000US-0229848P.
XX PR 01-SEP-2000; 2000US-0229850P.
XX PR 22-JAN-2001; 2001US-0263337P.
XX PR 31-JAN-2001; 2001US-0265518P.
XX PR 15-MAR-2001; 2001US-0276451P.
XX PR 27-MAR-2001; 2001US-0279196P.
XX PR 24-AUG-2001; 2001US-00939398.
XX PA (CURA-) CURAGEN CORP.
XX Gerlach V, Macdougall JR, Smithson G, Stone DJ, Ellerman K;
PI Spytek KA, Zedhusen BD, Rastelli L, Verney CAM, Patturajan M;
PI Tchernev VT, Padigaru M, Taupier RJ;
XX
XX WPI: 2002-292064/33.
XX N-PSDB; ARK55563.
XX
XX New isolated cytoplasmic, nuclear, membrane bound and secreted
PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune
PT disorders, haemophilia, Lesch-Nyhan syndrome, pancreatitis,
PT musculoskeletal disorders.
XX
XX Claim 1; Page 32; 245pp; English.
XX
XX The invention relates to an isolated cytoplasmic, nuclear, membrane bound
CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,
CC 4, 5a, 5b, 5c, 5d, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a
CC mature form, or a variant of the mature form of NOVX. Also included are a
CC polynucleotide encoding NOVX (or its complement), a vector comprising the
CC polynucleotide, a cell comprising the vector, an anti-NOVX antibody,
CC determining the presence of NOVX in a sample using the antibody,
CC which binds to NOVX polynucleotide, identifying an agent which binds to
CC NOVX (including modulators of NOVX). NOVX, the polynucleotide and the
CC antibody are useful for diagnosing, treating or preventing a NOVX-
CC associated disorder selected from cardiomyopathy, atherosclerosis,
CC diabetes, a disorder related to cell signal processing and metabolic
CC pathway modulation, inflammation, autoimmune disorders, scleroderma,
CC transplantation, allergies, systemic lupus erythematosus, haemophilia,
CC graft versus host disease, Alzheimer's disease, stroke, Lesch-Nyhan
CC syndrome, periodontitis, pancreatitis, musculoskeletal disorders,
CC Parkinson's disease, Huntington's disease, behavioural disorders, pain,
CC neurodegenerative and neuropsychiatric disorders, hypertension, wound
CC healing, obesity, growth and reproductive disorders, lung diseases and
CC many other diseases and disorders listed in the specification. NOVX, the
CC polynucleotide and the antibody are useful in screening assays, detection
CC assays (e.g., chromosomal mapping, tissue typing, forensic biology),
CC predictive medicine (e.g., diagnostic assays, prognostic assays,

CC monitoring clinical trials and pharmacogenomic), and in methods of
 CC treatment (e.g., therapeutic and prophylactic). NOVX is useful as
 CC immunogen to produce antibodies immunospecific for NOVX, as vaccines to
 CC screen for potential agonist and antagonist compounds, and as bait
 CC protein in a two-hybrid or three-hybrid assay. The polynucleotide is
 CC useful in gene therapy, to express NOVX, to detect NOVX mRNA or a genetic
 CC lesion in a NOVX gene, and to modulate NOVX activity. The vector is
 CC useful for producing non-human transgenic animals. The antibody is useful
 CC for isolating, and purifying NOVX and to monitor protein levels in tissue
 CC as part of a clinical testing procedure. The present sequence represents
 CC a NOVX protein
 XX
 SQ Sequence 1329 AA;

Query Match 41.1%; Score 3103.5; DB 5; Length 1329;
 Best Local Similarity 46.7%; Pred. No. 1.4e-130; Indels 321; Gaps 88;
 Matches 655; Conservative 0; Mismatches 426;
 1 GATCTGGGGTGTGCTCCAGGAAA-----AGCAAAATTCGGAA----- 37
 23 GCTCT-GCCAGCTACCTGGCAACGACACCCGCATCGCTGGTACCAACACCGCCCT 81
 38 GTTAATGTTTGTGATGATTTTAAATCTTGTGGGGAGAGCCCGCTC-TCCCGG 96
 82 GTGGAGGGTGTGAGC-AGGCGGGCATCTTC-CTGGCCGAGA-----GCCTCATCCACG- 133
 97 TATCAGCG-CTTCCTCATTTTGAATCCG-----GGCTCCCGGTCTTCGGCGT 146
 134 --ACTGCACCTTCATCACCAGGTGGCCCGCAACTCTGGCTGGCATCACACCTAC----- 186
 147 CAGACACCGGAGGAGCCCTTTTGAATTAAGC--GGCTGTGAACCCCGAGCGCG 204
 187 CAGTCTGCTGCTGATCTTCACTTCAGTGGCCCTTGGGCG-GGGGGTGCTCCCGGCGAC 245
 205 GGGGG-----GGCGGGCGGAGGCGGCGCTTTTGAATTAAGAGGCGGTGCTTCCA 255
 246 CCGAGCCTCCCGCGGTGTACCGT-GCGGCGCGTGGAGCGCAGGGGAC--TACTCCCA 302
 256 GCGAGGCTCTATA--AGTACCGCGCGGCGAGC-GTGCGC--GCTGTGAGTCTACTGT 310
 303 --CTGTCTCTACCAACAGCATACACAGGCTGTGTACACTTCG-TGCTGATGCCAT 359
 311 AGCGGACTTCTTTTGTCTTTCTTTCTTTCTTTGGGCGACT-CTGGAC-----TCACTCCC-- 363
 360 CAATGCTTCAATGCG-----CTGACC--TGCTCACCAGCTGGCGGTGTACACGCGA 413
 364 -----CAG-CATGAGGCGCTGAGCCCGT-GCGGGCGTGTAG----- 401
 414 GCGCGTAGCTTTTTCAGACATGATGATGTGTCTATGTGCTCAGATTCAGAAAT 473
 402 ---AGCGGTGTGCTGCTG-TCGGAACGCGAG-----TCTGCGCATCG 440
 474 TTTGGTTATGT-CGACCAGATCAAGAGCTGGTAGGGGTGATGGTGACATGGCCAGCA 532
 441 CCC-GGGCGCGAGGAGGCGCGCAGCTGAGGAGCGCTGAGCTGTGTGGAGCAGATG 499
 533 ACCTGATGCTGTGAGCAGCAC--CTGCTGT-----GCTGGCCGAGCGGAGGACAAG 585
 500 AACCACTGCTACT-----CC-----CGCC-TGCGGAACTGTATCCCGAGTCCC 543
 586 GCTGCGAGCGCATCGTGGTGGCTGTGAGGGCATTTGGGGGGCG-----CCCTCAGCCCC 642
 544 ---GAGAGGCA---CTCAGCTTA-GCCAG-----GTGGAAT-----CCTACAGCGGTCA 587
 643 CATGCCAGCATCTCAGTATGATGCGAGGAACTGGCATTTGAGGCGCTAC--CTATCA 700
 588 -----TCGACTACAT--TCTGACCTGCA-GGTAGTCTCTG-----GC 621
 701 AGCGGCACAGCTACGTGGCGCTGACCTGCACAGCGCTTCCAGAGGAGGAGGGGTGC 760
 622 CGAGCCAGC-CCCTGAGCCCGCTG-----ATGGCCCGCCACTT--TCCCAT---CCAGACAG 671
 761 CCGGCACACCGGCGGAGGAGCCTTGCCAGAACCCCGCCACTGTAGCCCGAGCCCGAGCTG 820

QY 672 CCGAGCTCGTCCGGAACTTGTCTCATCTCCAAACGACAAAGAGGAGCTTTTGCCACTG--ACT 729
 DB 821 ACCAGC-AGTCCG---CT--TCGCTGCACACACCGGGAGC-----CCATGTTTCT 867
 QY 730 CGGCGGTCTCTGACACCTTCAGAAC-GC-AGTGTCTGGGCGCGCTTCTGCTGGAGACC 787
 DB 868 CTGTC--GTCTTCTCCATCAAGAACAGCGTGGCCCTGCTCCATCCAGCT---CCCC 922
 QY 788 CGGG---AACTCT--CCTCCCGAAG-CGCGACGGCAGGATGGGCGCCCAACTTCGCC 841
 DB 923 CGAGTCTATTCTCATCTTCCGCTGCTGCTCCCGGTGCTCCCGGCTGCTCCCGCAGACTGCACC 982
 QY 842 TGC--CCACTTGACTTCAACCAAT-----CCCTTCTCT-----GGAGACTAAACCTGGTGC 889
 DB 983 TGCAGCTGCTGCTTTC-CGAATGGCGCTCTTCCACAGCAGCAACGCTCCCCG 1041
 QY 890 TCAGGAGCG--AAGGACTGTGAATTTGTAGCTTGAAGAG-----CCAG--AGCTAGCTC 939
 DB 1042 CTTGGAGCTGTGGGCTG---GCAAGAGGCGTGGGTGGCCACCGCTCATCTT-CGC 1097
 QY 940 TGG--GCACGAGCTG-GGCGACCTCACCTGCTCCACCCCGCCCGGCTTCTAAGGTC 996
 DB 1098 AGGAACAGTGTGTGTGGCTGGGAAACCTG-----ACAGTCACTGGCC 1143
 QY 997 TTTTCAGAGCG-----TGAGGTGTGAAGAGTGGCTGCT-----CTCCAAACT 1041
 DB 1144 GTTTCGCTGGCGCACTGGGCTGAGGAGCGCAACTGTGGCGCTTGTGGAGCCAGGAG 1203
 QY 1042 ATGCAAGGCGGGCG--AGAGCTGTCTTC--TGCTCTCTTGGAGAAGTCTGTGTTG 1097
 DB 1204 GGGCCCGG--GAGGCTGGGGCTGGACCTGGAGGGCTGCCAG--CTCGCTCCAGCA 1259
 QY 1098 CCTGTATTTATGAATCTATAAT-AGAGTATATAGTGTGTTGT-ACCTTTTACAGGAAG 1155
 DB 1260 GCGCAATGTACGCGCTGCACCTGCCAGCACTTGGGCAATGTGGCGTGTCT--CATGGAG 1317
 QY 1156 GTGACTTCTGTAACTATGCA 1177
 DB 1318 -----CTAGCCT---CGA 1328
 RESULT 9
 ABU8255
 ID ABU8255 standard; protein; 1743 AA.
 XX
 AC ABU8255;
 XX
 DT 07-JUL-2003 (first entry)
 XX
 DE Novel human secreted and transmembrane PRO polypeptide #3.
 XX
 KW Human; secreted and transmembrane protein: PRO; gene therapy;
 KW tumour necrosis factor-alpha release; TNF-alpha release;
 KW chondrocyte proliferation; chondrocyte differentiation; tumour;
 KW adrenal tumour; lung tumour; colon tumour; breast tumour;
 KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
 XX
 OS Homo sapiens.
 XX
 PN US2003032127-A1.
 XX
 PD 13-FEB-2003.
 XX
 PF 26-JUN-2002; 2002US-00183012.
 XX
 PR 18-SEP-1997; 97US-0059263P.
 PR 18-SEP-1997; 97US-0059266P.
 PR 17-OCT-1997; 97US-0062250P.
 PR 21-OCT-1997; 97US-0063486P.
 PR 24-OCT-1997; 97US-0063120P.
 PR 24-OCT-1997; 97US-0063121P.
 PR 28-OCT-1997; 97US-0063540P.

PR	28-OCT-1997;	97US-0063541P.	PR	11-JUN-1998;	98US-0088863P.
PR	28-OCT-1997;	97US-0063544P.	PR	11-JUN-1998;	98US-0088876P.
PR	28-OCT-1997;	97US-0063564P.	PR	12-JUN-1998;	98US-0089090P.
PR	29-OCT-1997;	97US-0063734P.	PR	12-JUN-1998;	98US-0089105P.
PR	31-OCT-1997;	97US-0063870P.	PR	16-JUN-1998;	98US-0089512P.
PR	31-OCT-1997;	97US-0064103P.	PR	16-JUN-1998;	98US-0089514P.
PR	13-NOV-1997;	97US-0065311P.	PR	17-JUN-1998;	98US-0089538P.
PR	21-NOV-1997;	97US-0066120P.	PR	17-JUN-1998;	98US-0089589P.
PR	24-NOV-1997;	97US-0066772P.	PR	17-JUN-1998;	98US-0089653P.
PR	11-DEC-1997;	97US-0069335P.	PR	18-JUN-1998;	98US-0089908P.
PR	12-DEC-1997;	97US-0069425P.	PR	19-JUN-1998;	98US-0089952P.
PR	17-DEC-1997;	97US-0069870P.	PR	22-JUN-1998;	98US-0090246P.
PR	18-DEC-1997;	97US-0068017P.	PR	22-JUN-1998;	98US-0090252P.
PR	10-MAR-1998;	98US-0077450P.	PR	22-JUN-1998;	98US-0090254P.
PR	11-MAR-1998;	98US-0077632P.	PR	24-JUN-1998;	98US-0090429P.
PR	11-MAR-1998;	98US-0077649P.	PR	24-JUN-1998;	98US-0090435P.
PR	20-MAR-1998;	98US-0078986P.	PR	24-JUN-1998;	98US-0090444P.
PR	20-MAR-1998;	98US-0078996P.	PR	24-JUN-1998;	98US-0090461P.
PR	27-MAR-1998;	98US-0079664P.	PR	24-JUN-1998;	98US-0090535P.
PR	27-MAR-1998;	98US-0079786P.	PR	24-JUN-1998;	98US-0090540P.
PR	31-MAR-1998;	98US-0080107P.	PR	25-JUN-1998;	98US-0090676P.
PR	01-APR-1998;	98US-0080337P.	PR	25-JUN-1998;	98US-0090686P.
PR	01-APR-1998;	98US-0081049P.	PR	25-JUN-1998;	98US-0090690P.
PR	08-APR-1998;	98US-0081070P.	PR	25-JUN-1998;	98US-0090694P.
PR	09-APR-1998;	98US-0081195P.	PR	25-JUN-1998;	98US-0090696P.
PR	15-APR-1998;	98US-0082568P.	PR	25-JUN-1998;	98US-0090699P.
PR	21-APR-1998;	98US-0082569P.	PR	26-JUN-1998;	98US-0090862P.
PR	21-APR-1998;	98US-0082704P.	PR	26-JUN-1998;	98US-0090863P.
PR	22-APR-1998;	98US-0082797P.	PR	26-JUN-1998;	98US-0091010P.
PR	28-APR-1998;	98US-0083322P.	PR	01-JUL-1998;	98US-0091359P.
PR	29-APR-1998;	98US-0083496P.	PR	01-JUL-1998;	98US-0091544P.
PR	29-APR-1998;	98US-0083499P.	PR	02-JUL-1998;	98US-0091478P.
PR	29-APR-1998;	98US-0083499P.	PR	02-JUL-1998;	98US-0091486P.
PR	29-APR-1998;	98US-0084366P.	PR	02-JUL-1998;	98US-0091626P.
PR	05-MAY-1998;	98US-0084414P.	PR	02-JUL-1998;	98US-0091628P.
PR	07-MAY-1998;	98US-0084639P.	PR	02-JUL-1998;	98US-0091632P.
PR	07-MAY-1998;	98US-0084640P.	PR	02-JUL-1998;	98US-0094006P.
PR	17-MAY-1998;	98US-0084643P.	PR	04-AUG-1998;	98US-0095282P.
PR	15-MAY-1998;	98US-0085579P.	PR	10-AUG-1998;	98US-0095998P.
PR	15-MAY-1998;	98US-0085580P.	PR	10-AUG-1998;	98US-0096012P.
PR	15-MAY-1998;	98US-0085582P.	PR	17-AUG-1998;	98US-0096757P.
PR	15-MAY-1998;	98US-0085700P.	PR	17-AUG-1998;	98US-0096765P.
PR	18-MAY-1998;	98US-0086023P.	PR	17-AUG-1998;	98US-0096867P.
PR	22-MAY-1998;	98US-0086392P.	PR	17-AUG-1998;	98US-0096891P.
PR	28-MAY-1998;	98US-0086486P.	PR	17-AUG-1998;	98US-0096897P.
PR	28-MAY-1998;	98US-0087098P.	PR	18-AUG-1998;	98US-0096949P.
PR	28-MAY-1998;	98US-0087208P.	PR	18-AUG-1998;	98US-0096959P.
PR	02-JUN-1998;	98US-0087609P.	PR	18-AUG-1998;	98US-0097022P.
PR	02-JUN-1998;	98US-0087759P.	PR	26-AUG-1998;	98US-0097952P.
PR	04-JUN-1998;	98US-0088025P.	PR	26-AUG-1998;	98US-0097955P.
PR	04-JUN-1998;	98US-0088028P.	PR	26-AUG-1998;	98US-0097971P.
PR	04-JUN-1998;	98US-0088029P.	PR	26-AUG-1998;	98US-0097974P.
PR	04-JUN-1998;	98US-0088033P.	PR	26-AUG-1998;	98US-0098014P.
PR	04-JUN-1998;	98US-0088326P.	PR	01-SEP-1998;	98US-0098716P.
PR	05-JUN-1998;	98US-0088167P.	PR	01-SEP-1998;	98US-0098723P.
PR	05-JUN-1998;	98US-0088202P.	PR	02-SEP-1998;	98US-0098803P.
PR	05-JUN-1998;	98US-0088212P.	PR	02-SEP-1998;	98US-0098821P.
PR	05-JUN-1998;	98US-0088217P.	PR	02-SEP-1998;	98US-0098843P.
PR	09-JUN-1998;	98US-0088655P.	PR	03-SEP-1998;	98US-0099602P.
PR	09-JUN-1998;	98US-0088722P.	PR	10-SEP-1998;	98US-0099741P.
PR	10-JUN-1998;	98US-0088738P.	PR	10-SEP-1998;	98US-0099754P.
PR	10-JUN-1998;	98US-0088740P.	PR	10-SEP-1998;	98US-0099763P.
PR	10-JUN-1998;	98US-0088811P.	PR	10-SEP-1998;	98US-0099812P.
PR	10-JUN-1998;	98US-0088824P.	PR	15-SEP-1998;	98US-0100388P.
PR	10-JUN-1998;	98US-0088825P.	PR	16-SEP-1998;	98US-0100662P.
PR	10-JUN-1998;	98US-0088826P.	PR	16-SEP-1998;	98US-0100664P.
PR	11-JUN-1998;	98US-0088861P.	PR	16-SEP-1998;	98US-0101751P.
PR			PR	16-SEP-1998;	98US-0101933P.
PR			PR	16-SEP-1998;	98US-0100683P.
PR			PR	17-SEP-1998;	98US-0100684P.

PR	17-SEP-1998;	98US-0100919P.	Query Match	41.1%;	Score 3101.5;	DB 6;	Length 1743;
PR	17-SEP-1998;	98US-0100930P.	Best Local Similarity	41.8%;	Pred. No. 2.1e-130;		
PR	18-SEP-1998;	98US-0100849P.	Matches	725;	Conservative	0;	Mismatches 421; Indels 587; Gaps 102;
PR	18-SEP-1998;	98US-0101014P.					
PR	18-SEP-1998;	98US-0101068P.					
PR	23-SEP-1998;	98US-0101471P.					
PR	23-SEP-1998;	98US-0101472P.					
PR	23-SEP-1998;	98US-0101475P.					
PR	23-SEP-1998;	98US-0101477P.					
PR	24-SEP-1998;	98US-0101738P.					
PR	24-SEP-1998;	98US-0101739P.					
PR	24-SEP-1998;	98US-0101743P.					
PR	24-SEP-1998;	98US-0101922P.					
PR	25-SEP-1998;	98US-0101786P.					
PR	25-SEP-1998;	98US-0102207P.					
PR	29-SEP-1998;	98US-0102240P.					
PR	29-SEP-1998;	98US-0102330P.					
PR	29-SEP-1998;	98US-0102331P.					
PR	30-SEP-1998;	98US-0102487P.					
PR	30-SEP-1998;	98US-0102570P.					
PR	30-SEP-1998;	98US-0102571P.					
PR	01-OCT-1998;	98US-0102684P.					
PR	01-OCT-1998;	98US-0102687P.					
PR	02-OCT-1998;	98US-0102965P.					
PR	06-OCT-1998;	98US-0103258P.					
QY	1	GATCTGGG-----GTGCTGCCAGGAAAGCAAAATCTGGAAGTAAATGTTTGGAG 52					
DB	12	GGACAGGCGAGGAGAGGAGGAACTGCGAAGA-CATATTTGTTCCAAAATGGCATC--- 67					
QY	53	TGATTTT-----AATCTTCTGCGGAGAGGCGCCCTCT---CCCGGTAT-CAG 102					
DB	68	TTACCTTTATGGAGTACTCTTTGCTGTG-----GCCCTGTGCTCCAACTACTG 118					
QY	103	CGCTTCCTCAATCTTTGAATCC---GGGGCTCCGGGTCTTTCGGCGTC---AGACACAGCG 157					
DB	119	TGTTCTCCCGGCAATGCCCCAGTGCATACCCCGCCCTTCTCCACAAAGAGCACCC- 177					
QY	158	GAGNAACC-----TGTTTGCATTTAAGCGGGTGTG---AAGCCCCAGG--GCCGG 205					
DB	178	-----CTGCCCTCACAGGTATTCCTTCAACACCGACTTTGCTTCCGCTATACCCGAG 233					
QY	206	CGGG-----GGCGGGCGGCGCGGCGCATTT-----TGAATAAAGAGGCGTGC 249					
DB	234	CTGGTTTGGAGACCCCGAGTCAGACATCTTCTTCTCCCTGTGAGTGTCTCCACTCC 293					
QY	250	CT-TCCAGG-----AGGCTCTATAAGTGACC-----GCC-- 278					
DB	294	CTGGCCATGTCTCTCCCTTGGGGCCCACTCAGTCACCAAGACCCAGATTTCTCAGGGCC 353					
QY	279	-GGCGGAGCGTGGCGC-----GT-TGCAGGTCACTGTAGCGGACTTCTTTTGGTT 328					
DB	354	GGCTTCAACCTCACACACACAGAGTCTGCCATCCACCAGGCTTCCAGCACCTGGTT 413					
QY	329	TTCTTTCTCTTTGG-----GGC-----ACCT---CTGGAC-----TCACCTCCCCAGC 367					
DB	414	CACCTCACTGACTGTTCCTCCAGCAAGACCTGACCTTGAAGATGGGAAAGTGCCCTCTTCGTC 473					
QY	368	ATGAAGCGCTGAGCCCGTGC-----GCC-----TGCTACGAGCGGTGTGCT 413					
DB	474	AAGAAGAGCTG-----CAGCTGCAGGCAAAATTTCTTGGCAATGTCAAGAGCTGTATGA 530					
QY	414	GC-----CTGT-----CGGA-----ACGAGTGGCCATCGCCCGGGCGCGAGG---AAG 457					
DB	531	GCAGAAGTCTTTTCTACAGATTTCTC-CAACCCCTCCATTGGCCAGG--CGAGGATCAAC 587					
QY	458	GGCCCGGAGCTGAGGAGCCCTGAGTGTGTGACGACATGACCACTGC-----TACTC 513					
DB	588	AGCCATGTGAAGAGAGACCAAGGAGGTTGTAGACATATCCAGGSCCTTGACCTT 647					

QY	514	CCGCTGC--GGGAACGTGGT--ACCGGAGTC-----CC-----GAGAGGCAC--T 553
DB	648	CTGACGCCCATGGTTCTGTGTAATCACAATTTCTTTAAAGCCAAAGTGGGAGAGCCCTTT 707
QY	554	CAGCTT-----AG-----CCAGGTGG-----AA 571
DB	708	CACCTTGAATATACAGAAAGAACTTCCCATTTCTGTGGGCGAGAGGTCACTGTGCAA 767
QY	572	ATCC-----TACAG--CGCGT-----CATCG-----ACTAC 595
DB	768	GTCCCATGATGCACCAAGAGAGCAGTTCGCTTTTGGGGTGGATACAGAGCTGAATGC 827
QY	596	ATTCTCGACCTGCAGGTAG--TCCTGGCGGA--GCC--AGCC--CCTGGACC--CCCT-- 643
DB	828	TTT---GTGTGCGAGATGATTAAGGGAGATGCGGTGCGCTCTTCTTGTCTCCCTAGC 884
QY	644	-----GAT--GGCCCC-----CA--CCTTCCATCCAGACAGCCGA----- 675
DB	885	AAGGCGCAAGATGAGGCAACTGGAACAGGCTTGTGACGCAAGACACTGATAAAGTGAGC 944
QY	676	-GCTCGCTCCGGACTTG-----TCATCTCCACGACAAAGAGGCTTTTC 721
DB	945	CACCTCACTCCAGAAAGGTGGATAGAGTGTTCATCCC-----AGATTTT-C 991
QY	722	CA---CTGACTCGGCGCTGTCTGACACCTCCAGAACGCGAGGTCTGGCGCCGCTTCTGC 778
DB	992	CATTTCTGCTCTTAC--AATCTGGAACCATCTCTCCGAAAGATG--GGCATCCAAAATGC 1048
QY	779	CTGGGACC-----CCGG-----GAACCTCTCTCTCCCGG----- 806
DB	1049	CTTTGACAAAATCTGATTTTCTGGAATTCMAAGAGAGACTC-CTGCAGTCTCTA 1107
QY	807	---AAGCGGACGGCAG--GGATG-----GGCC----- 829
DB	1108	AAGCAACCCACAGCTGTGCTGATGTCAAGAGAGGCGCACTGAGGCCACAGCAGTA 1167
QY	830	---CCAACTTC--GCC-----TG--CCCACTTGACTTCCCAATTCCTT 868
DB	1168	CCACCAACCAAGTTCATAGTCGATCGAAGAGTGTCTCTT-ACTTCACTGTCTCTTC 1226
QY	869	CCTGAGACTAAACCTGTGTCT-----CA---GGAGCAAGGAGTGTGAAT----- 912
DB	1227	AATAGGACCT--TCCTGATGATGATTACAAATAAGCCACAGAGGTATTCTCTTTCTAG 1284
QY	913	---TGTAAGAGCCAGAGTGTCTGGC-----ACCAGCTGGGCGAC 958
DB	1285	GGAAAGTGGAAATCCCACTAAATCTAGTGGGAAATGGCTGTAACTGATG----- 1339
QY	959	GTCACCTGTCT-----CCCAACCCACCC-----CAAGTTCTAAGG--TC 996
DB	1340	---CATTGCTAATGCACAGAAATAACCAACATCCCTCTTCTGTCTGAGGTGC 1397
QY	987	TTTTTCAAGC--GTGGAGTG-----TGAAGAGGTGGCT-----GC-----TCTCC 1036
DB	1398	ATTTGACCCAGTGGAGCTGGATTTCGCTGGCAGGATGCCACTTCCAAGGCTCAATCACC 1457
QY	1037	AAACTAT-----GCCAAGGC-----GGC-----GGCAGAG--CTG 1064
DB	1458	AAACATCAACAGGAGCCCGAGTCACAGCCCAACCCATTAACCCAGTCAGTGCCTT 1517
QY	1065	GTCTTCTGG-TCTCCTTGG-AGAAAGGTTC-----TGTTC-----CCTGATTT--A 1107
DB	1518	TTCCACAAATTTCTCCAGGTAACTAGTTCATGATGGATGTGCTGGGTACCATATTCCA 1577
QY	1108	T-----GAACCTATAATAGAG-----TATATAGGTTTGTACCTTTTACAGG 1152
DB	1578	TTCTTTGGGGTCCAGGAATGGAATAACGCCAACCCAGGTAGGACCTCTATTGAGA 1637
QY	1153	-----AAGGTGACTTTCTGTAAACAATCGATGTATATTAATTTTATAAAA 1200
DB	1638	ATTACATAACACATTTCAATAAATAAATAATGAATTCAAAAAATAAAAAA 1690

RESULT 10
ABU90134
ID ABU90134 standard; protein; 1743 AA.
XX AC
XX ABU90134;
XX DT
XX 11-AUG-2003 (first entry)
XX DE
XX DE Novel human secreted and transmembrane PRO protein #3.
XX DE
XX DE Human; gene therapy; tissue typing; tumour; chondrocyte proliferation;
XX KW chondrocyte differentiation; tumour necrosis factor-alpha release;
XX KW affinity purification.
XX OS
XX OS Homo sapiens.
XX PN
XX PN US2003036147-A1.
XX PD
XX PD 20-FEB-2003.
XX PF
XX PF 02-JUL-2002; 2002US-00187741.
XX PR
XX PR 18-SEP-1997; 97US-0059263P.
XX PR 18-SEP-1997; 97US-0059266P.
XX PR 17-OCT-1997; 97US-0062250P.
XX PR 21-OCT-1997; 97US-0063486P.
XX PR 24-OCT-1997; 97US-0063120P.
XX PR 24-OCT-1997; 97US-0063121P.
XX PR 28-OCT-1997; 97US-0063540P.
XX PR 28-OCT-1997; 97US-0063541P.
XX PR 28-OCT-1997; 97US-0063544P.
XX PR 28-OCT-1997; 97US-0063564P.
XX PR 29-OCT-1997; 97US-0063734P.
XX PR 31-OCT-1997; 97US-0063870P.
XX PR 31-OCT-1997; 97US-0064103P.
XX PR 13-NOV-1997; 97US-0065311P.
XX PR 21-NOV-1997; 97US-0066120P.
XX PR 24-NOV-1997; 97US-0066466P.
XX PR 11-DEC-1997; 97US-0066772P.
XX PR 12-DEC-1997; 97US-0068335P.
XX PR 17-DEC-1997; 97US-0069425P.
XX PR 18-DEC-1997; 97US-0069870P.
XX PR 10-MAR-1998; 98US-0077450P.
XX PR 11-MAR-1998; 98US-0077632P.
XX PR 11-MAR-1998; 98US-0077649P.
XX PR 20-MAR-1998; 98US-0078896P.
XX PR 20-MAR-1998; 98US-0078939P.
XX PR 27-MAR-1998; 98US-0079664P.
XX PR 27-MAR-1998; 98US-0079786P.
XX PR 31-MAR-1998; 98US-0080107P.
XX PR 31-MAR-1998; 98US-0080194P.
XX PR 01-APR-1998; 98US-0080333P.
XX PR 08-APR-1998; 98US-0081049P.
XX PR 08-APR-1998; 98US-0081070P.
XX PR 09-APR-1998; 98US-0081195P.
XX PR 15-APR-1998; 98US-0081838P.
XX PR 21-APR-1998; 98US-0082568P.
XX PR 21-APR-1998; 98US-0082569P.
XX PR 22-APR-1998; 98US-0082704P.
XX PR 22-APR-1998; 98US-0082797P.
XX PR 28-APR-1998; 98US-0083322P.
XX PR 29-APR-1998; 98US-0083495P.
XX PR 29-APR-1998; 98US-0083496P.
XX PR 29-APR-1998; 98US-0083499P.
XX PR 05-MAY-1998; 98US-0083559P.
XX PR 05-MAY-1998; 98US-0084366P.
XX PR 06-MAY-1998; 98US-0084414P.
XX PR 07-MAY-1998; 98US-0084633P.
XX PR 07-MAY-1998; 98US-0084640P.
XX PR 07-MAY-1998; 98US-0084643P.

PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086496P.
PR 28-MAY-1998; 98US-0087038P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088039P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089514P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090423P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090675P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 26-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0090866P.
PR 01-JUL-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.

```
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098169P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 13-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-01019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 25-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-0016897P.

PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098169P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 13-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-01019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 25-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-0016897P.

Query Match 41.1%; Score 3101.5; DB 6; Length 1743;
Best Local Similarity 41.8%; Pred. No. 2.1e-130;
Matches 725; Conservative 0; Mismatches 421; Indels 587; Gaps 102;

QY 1 GATCTGGG-----GTGCTGCAGGAAAGCAATTCGGAAGTTAATGTTTGTAG 52
DB 12 GGACAGGCGGAGAGAGAGACCTTGCAGAG-CATATTTTGTCCAAATGGCATC--- 67
QY 53 TGATTTTAA-----AATCTTGTGGCGGAGAGCGCCGCTCT---CCCGGATP-CAG 102
DB 68 TTACCTTTATGGAGTACTTTTGTCTGTTG-----GCCTGTGTCTCAATCTACTG 118
QY 103 CGCTTCTCATCTTTTGAATCC---CGGCTCCGGGCTTTGCGGCTC--AGACCGCG 157
DB 119 TGTGTCCCGGCAATGCCCCAGTGCATACCCCGCCCTTCTCCACAGAGACCCC- 177

QY 158 GAGGAAGCC-----TGTTTGAATTAAGCGGGCTGTG---AACCCGAGG--GCCG 205
DB 178 ----CTGGCTCACAGGTGTATTCCTCAACACGACTTTCCTTCCGCTTATACCGCAGG 233
QY 206 CGGG-----GGCGGGCCGAGCGCGGCCATTT-----TGAATAAGAGCGCTGC 249
DB 234 CTGGTTTGGAGACCCCGAGTCAGAACATCTTCTTCTCCCTGTGAGTGTCTCCACTTC 293
QY 250 CT-TCCAGGC-----AGGCTCTATAAGTGACC-----GGC-- 278
DB 294 CTGGCCATGCTCTCCCTTGGGGCCCACTCACTCACCAGACCCAGATCTCCAGGGCTG 353
QY 279 -GGCGGAGCGTGGCGC-----GT-TGAGGTCACTGTAGCGGACTCTTTTGGTT 328
DB 354 GGCCTTCAACCTCACACACACAGAGCTGCCATCCACGAGGCTTCCAGGACCTGGTT 413
QY 329 TTTCTTCTCTTTGG-----GGC-----ACCT---CTGGAC-----TCACTCCCCAGC 367
DB 414 CACTCACTGACTGTTTCCAGCAAGACCTGACCTTGAAGATGGGAAGTGCCTTCTGTC 473
QY 368 ATGAAGCGCTGAGCCCGGTG-----GGC-----GGC-TGCTACGAGCGGTGTGCT 413
DB 474 AAGAAGGAGCTG---CAGCTCAGGCAATTTCTGGCAATGTCAAGAGCGTGTATGAA 530
QY 414 GC-----CTGT-----CGGA-----ACGAGTCTGCCATCGCCGGGCGGAGG--AAG 457
DB 531 GCAGAGTCTTTTCTACAGATTTCTC-CAACCCCTCATTTGCCAGG--CGAGATCAAC 587
QY 458 GCGCCGCGAGCTGAGGCGCTGAGCTTCTGACGACGATGAACACTGC-----TACTC 513
DB 588 AGCCATGTGAAAAGAGACCCCAAGGAGGTTGTAGACATAATCCAAAGCCTTGACCTT 647
QY 514 CGGCTGC--GGGAACGTGT---ACCCGGAGTC-----CC-----GAGAGGCAC--T 553
DB 648 CTGACGGCCATGTTCTGTTGAATCAATTTCTTTAAAGCAAGTGGGAGAGCCCTTT 707
QY 554 CAGCTT-----AG-----CGAGTGG-----AA 571
DB 708 CACTTGAATATACAGAAAGAACTTCCCATCTGTGTGGCGAGAGGTCACGTGTGCA 767
QY 572 ATCC-----TACAG--CGCCT-----CATCG-----ACTAC 595
DB 768 GTCCCATGATGCACCAAGAGACGCTTCTGCTTTTGGGTGGATACAGAGTGAAGTGC 827
QY 596 ATCTCTGACCTGCAGGTAG--TCTGCGCA-GCC--AGCC--CCTGGAC--CCCT-- 643
DB 828 TTT---GTGCTGCAGATGGATTACAGGAGATGCCGTGCTTCTTTGCTCCCTAGC 884
QY 644 -----GAT--GGCCCC-----CA--CCTTCCCATCCACAGCCGA----- 675
DB 885 AAGGCAAGATGAGGCAACTGGAACAGGCTTGTGAGCCAGAACACTGATAAAGTGGAGC 944
QY 676 -GCTCGCTCCGGAACCTTG-----TCACTCTCAACGACAAAAGAGCTTTTGC 721
DB 945 CACTCACTCCAGAAAGGTGATAGAGTGTTCATCCCC-----AGATTTT-C 991
QY 722 CA---CTGACTCGGCGCTGTCTGACACTTCCAGAGCGAGGTGTGCGGCCCGCTTCTGC 778
DB 992 CATTTCTGCTCTCTAC-AACTGGAACCATCTCCCGAGATG--GGCATCCAAAATGC 1048
QY 779 CTGGAGC-----CCGG-----GAACCTCTCTCTGCGG----- 806
DB 1049 CTTTGACAAATGCTGATTTTCTGAAATGCAAGAGAGACTC-CTTGCAGGTTCTTA 1107
QY 807 ----AAGCCGACCGCAG-----GGATG-----GGCC----- 829
DB 1108 AAGCAACCCACAGGCTGTGCTGGATGTGAGTGAAGAGGCGCACTGAGGCGACAGACTA 1167
QY 830 -----CCAATTC--GCCC-----TG--CCACTTCACTTCACTCAATCCCTT 868
DB 1168 CCACCAACAAGTTTATAGTCGATCGATCGAAGGATGTCCTCTT-ACTTCACTGTCTCTC 1226
```

QY 869 CTGGAGCTAAACCTGGTCT-----CA---GGAGCGAAGGAGCTGGAAC----- 912
 Db 1227 AATAGGACCT--TCTGATGATGATTAACAATAAAGCCAGACGGTATCTCTTTCTAG 1284
 QY 913 ----TGTAAG-----CCTGAAGAGCCAGAGCTAGCTCTGGCC-----ACCAGCTGGCGGAC 958
 Db 1285 GGAAGTGGAAATCCCACTAAATCTTAGTGGGAATGCGCTGTTAACTGATGG----- 1339
 QY 959 GTACACCTGCT-----CCACCCCAACCC-----CAAGTTCTAAGG--TC 996
 Db 1340 --CACATTGCTAATGCAAGAATAACAACCAACCATCCCTCTTTCTGTTCTGAGGGTGC 1397
 QY 997 TTTTCAGAGC-GTGAGGTG-----TGGAAGGAGTGGGT-----GC-----TCTCC 1036
 Db 1398 ATTTGACCCAGTGGAGTGGATTCGCTGGCAGGATGCCACITCCCAAGGCTCAATCAAC 1457
 QY 1037 AAATAT-----GCCAAGC-----GGC-----GCCAGAG--CTG 1064
 Db 1458 AAACCATCAACAGGAGCCCAAGTCCACAGCCCAACCCCAATTAACCCAGTCAGTGCCTT 1517
 QY 1065 GTCTTCTGG-TCTCCTGG-AGAAAGTTC-----TGTTGC-----CCTGATTT--A 1107
 Db 1518 TTCACAAATTTCTCCAGTAACTAGTCTCATGGATGTTGCTGGTTACCATATTTCCA 1577
 QY 1108 T-----GAACTCTATAATAGAG-----TATATAGTTTGTACCTTTTATACAGG 1152
 Db 1578 TTCCTTGGGGTCCAGGAATGGAATACGCCAACCCAGGTTAGGCACCTCTATTGCAGA 1637
 QY 1153 -----ARGTGACTTCTGTACCAATGCGATGATATTAACCTTTTATAAAA 1200
 Db 1638 ATTACAATAACACATTCATAAATAAATGAATGATCAAAAAAATAA 1690

RESULT 11
 ABU96436
 ID ABU96436 standard; protein; 1743 AA.
 AC ABU96436;
 DT 25-JUL-2003 (first entry)
 DE Novel human secreted and transmembrane protein #3.
 KW Human; secreted and transmembrane protein; PRO; transgenic animal;
 KW knockout; chromosome identification; tissue typing; tumour;
 KW chondrocyte proliferation; chondrocyte differentiation;
 KW tumor necrosis factor-alpha release stimulator.
 OS Homo sapiens.
 XX
 XX US2003036144-A1.
 PD 20-FEB-2003.
 XX
 XX 01-JUL-2002; 2002US-00187601.
 XX
 XX 18-SEP-1997; 97US-0059263P.
 XX 18-SEP-1997; 97US-0059266P.
 XX 17-OCT-1997; 97US-0062250P.
 XX 21-OCT-1997; 97US-0063486P.
 XX 24-OCT-1997; 97US-0063120P.
 XX 24-OCT-1997; 97US-0063121P.
 XX 28-OCT-1997; 97US-0063540P.
 XX 28-OCT-1997; 97US-0063541P.
 XX 28-OCT-1997; 97US-0063544P.
 XX 28-OCT-1997; 97US-0063564P.
 XX 29-OCT-1997; 97US-0063734P.
 XX 31-OCT-1997; 97US-0063870P.
 XX 31-OCT-1997; 97US-0064103P.
 XX 13-NOV-1997; 97US-0065311P.
 XX 21-NOV-1997; 97US-0066120P.
 XX 24-NOV-1997; 97US-0066466P.
 XX 24-NOV-1997; 97US-0066772P.

PR 11-DEC-1997; 97US-0069335P.
 PR 12-DEC-1997; 97US-0069425P.
 PR 17-DEC-1997; 97US-0069870P.
 PR 18-DEC-1997; 97US-0068017P.
 PR 10-MAR-1998; 98US-0077450P.
 PR 11-MAR-1998; 98US-0077632P.
 PR 11-MAR-1998; 98US-0077649P.
 PR 20-MAR-1998; 98US-0078866P.
 PR 20-MAR-1998; 98US-0078939P.
 PR 27-MAR-1998; 98US-0079664P.
 PR 27-MAR-1998; 98US-0079786P.
 PR 31-MAR-1998; 98US-0080107P.
 PR 31-MAR-1998; 98US-0080194P.
 PR 01-APR-1998; 98US-0080327P.
 PR 01-APR-1998; 98US-0080333P.
 PR 08-APR-1998; 98US-0081049P.
 PR 08-APR-1998; 98US-0081070P.
 PR 09-APR-1998; 98US-0081195P.
 PR 15-APR-1998; 98US-0081838P.
 PR 21-APR-1998; 98US-0082568P.
 PR 21-APR-1998; 98US-0082569P.
 PR 22-APR-1998; 98US-0082704P.
 PR 22-APR-1998; 98US-0082797P.
 PR 28-APR-1998; 98US-0083322P.
 PR 29-APR-1998; 98US-0083495P.
 PR 29-APR-1998; 98US-0083496P.
 PR 29-APR-1998; 98US-0083499P.
 PR 29-APR-1998; 98US-0083559P.
 PR 05-MAY-1998; 98US-0084366P.
 PR 06-MAY-1998; 98US-0084414P.
 PR 07-MAY-1998; 98US-0084639P.
 PR 07-MAY-1998; 98US-0084640P.
 PR 07-MAY-1998; 98US-0084643P.
 PR 15-MAY-1998; 98US-0085579P.
 PR 15-MAY-1998; 98US-0085580P.
 PR 15-MAY-1998; 98US-0085582P.
 PR 15-MAY-1998; 98US-0085700P.
 PR 18-MAY-1998; 98US-0086023P.
 PR 22-MAY-1998; 98US-0086392P.
 PR 22-MAY-1998; 98US-0086486P.
 PR 28-MAY-1998; 98US-0087098P.
 PR 28-MAY-1998; 98US-0087208P.
 PR 02-JUN-1998; 98US-0087609P.
 PR 02-JUN-1998; 98US-0087759P.
 PR 03-JUN-1998; 98US-0087827P.
 PR 04-JUN-1998; 98US-0088025P.
 PR 04-JUN-1998; 98US-0088028P.
 PR 04-JUN-1998; 98US-0088029P.
 PR 04-JUN-1998; 98US-0088033P.
 PR 05-JUN-1998; 98US-0088326P.
 PR 05-JUN-1998; 98US-0088167P.
 PR 05-JUN-1998; 98US-0088202P.
 PR 05-JUN-1998; 98US-0088212P.
 PR 05-JUN-1998; 98US-0088217P.
 PR 09-JUN-1998; 98US-0088655P.
 PR 10-JUN-1998; 98US-0088722P.
 PR 10-JUN-1998; 98US-0088738P.
 PR 10-JUN-1998; 98US-0088740P.
 PR 10-JUN-1998; 98US-0088811P.
 PR 10-JUN-1998; 98US-0088824P.
 PR 10-JUN-1998; 98US-0088825P.
 PR 10-JUN-1998; 98US-0088826P.
 PR 11-JUN-1998; 98US-0088861P.
 PR 11-JUN-1998; 98US-0088863P.
 PR 11-JUN-1998; 98US-0088876P.
 PR 12-JUN-1998; 98US-0089090P.
 PR 12-JUN-1998; 98US-0089105P.
 PR 16-JUN-1998; 98US-0089512P.
 PR 16-JUN-1998; 98US-0089514P.
 PR 17-JUN-1998; 98US-0089538P.
 PR 17-JUN-1998; 98US-0089598P.
 PR 17-JUN-1998; 98US-0089653P.
 PR 18-JUN-1998; 98US-0089908P.

PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 02-OCT-1998; 98US-0102687P.
PR 06-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.

Query Match 41.1%; Score 3101.5; DB 6; Length 1743;
Best Local Similarity 41.8%; Pred. No. 2.1e-130;
Matches 725; Conservative 0; Mismatches 421; Indels 587; Gaps 102;

QY 1 GATCTGG-----GTGCTGCCAGAAAAGCAAAATCTGGAAGTTAATGTTTTGAG 52
DB 12 GGAAGGAGGAGAGAGAACTGCAAGA-CATATTTTGTCCAAATGGCATC--- 67

QY 53 TGATTTTA-----AATCCTTGCTGGCGAGAGCCGCCCTCT---CCCCGGTAT-CAG 102
DB 68 TTACCTTTATGGAGTACTCTTTGCTGTG-----GCCCTGTGCTCCAACTACTG 118

QY 103 CGCTTCTCATTTCTTTGAATCC---CGGCTCGGGGTCTTCGGGTC--AGACAGCGG 157
DB 119 TGTGTCCTCCGCGCAATGCCCCAGTGCATACCCCCCTTCTCCAAAAGAGACCC- 177

QY 158 GAGGAAGCC-----TGTTTGCAATTTAAGCGGGCTGTG---AACGCCAGG--GCCGG 205
DB 178 ----CTGCTTCACAGGTGATTCCTCAACACCGACTTTGCCTTCGCTATACCCGAG 233

QY 206 CGGG-----GGCGGGCGGAGCGGGCCATTT-----TGATAAGAGGGGTGC 249
DB 234 CTGGTTTGGAGACCCCGAGTCAGAACATCTTCTTCCCTGTGAGTGTCTCCACTTC 293

QY 250 CT-TCCAGGC-----AGGCTCTATAAGTGACC-----GCC-- 278
DB 294 CTGGCCATGCTCTCCCTTGGGGCCCACTCAGTCAACAGACCCAGATTCTCCAGGGCTG 353

QY 279 -GGCGGAGCGTGGCGC-----GT-TGCAGGTCACTGTAGCGGACTTCTTTGGTT 328
DB 354 GGCTTCAACCTCACACACACAGAGTCTGCCATCCACAGGGCTTCAGACCTGGTT 413

QY 329 TTCTTTCTCTTTGG---GGC-----ACCT---CTGGAC-----TCACTCCCCAGC 367
DB 414 CACTCACTGACTGTCCAGAAAGACCTGACCTTGAAGATGGAAGTGCCTCTTCGTC 473

QY 368 ATGAAGCGCTGAGCCCGGTGC--GC-----GGC--TCCTAGAGGGGGTGTGCT 413
DB 474 AAGAAGGAGCTG---CAGCTGCAAGGCAAAATTTCTTGGGCAATGCAAGAGGCTGTATGA 530

QY 414 GC-----CTGT-----CGGA---AGCGAGTCTGGCCATCGCCGGGGCGGAGG--AAG 457
DB 531 GCAGAGTCTTTTCTACAGATTCTC-CAACCCCTCAATGCCCAGG--CGAGGATCAAC 587

QY 458 GGCCCGGACCTGAGAGCGCGTGTGCTGTGGAGACATGTAACCACTGC-----TACTC 513
DB 588 AGCCATGTGAAAAGGAAGACCCAAAGGAGGTTGTAGACATAATCCAAGGCTTTGACCT 647

QY 514 CCGCTGC--GGGAACCTGGT---ACCGGAGTC-----CC-----GAGAGGCAC--T 553
DB 648 CTGACGCCCATGGTTCTGGTGAATCACAATTTCTTTAAAGCCAAAGTGGGAGAGCCCTTT 707

QY 554 CAGCTT-----AG-----CAAGGTG-----AA 571
DB 708 CACCTTGATATATAAGAAAGAACTTCCCATTCCTGTGGCGAGCAGGTCACTGTGCAA 767

PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090433P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 03-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 10-SEP-1998; 98US-0100388P.
PR 10-SEP-1998; 98US-0100662P.
PR 10-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-01019330.
PR 16-SEP-1998; 98US-0101751P.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.

Novel human secreted and transmembrane protein #3.

Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy; chondrocyte stimulator; tumour; adrenal tumour; lung tumour; colon tumour; breast tumour; prostate tumour; rectal tumour; cervical tumour; liver tumour; TNF-alpha release; tumour necrosis factor alpha release; chondrocyte cell proliferation; chondrocyte cell differentiation; pharmaceutical; diagnostic; biosensor; bioreactor.

Homo sapiens.

US2003013153-A1.

16-JAN-2003.

19-JUN-2002; 2002US-00175737.

18-SEP-1997; 97US-0059263P.
18-SEP-1997; 97US-0059266P.
17-OCT-1997; 97US-0062250P.
21-OCT-1997; 97US-0063486P.
24-OCT-1997; 97US-0063120P.
24-OCT-1997; 97US-0063121P.
28-OCT-1997; 97US-0063540P.
28-OCT-1997; 97US-0063541P.
28-OCT-1997; 97US-0063544P.
28-OCT-1997; 97US-0063564P.
29-OCT-1997; 97US-0063734P.
31-OCT-1997; 97US-0063870P.
31-OCT-1997; 97US-0064103P.
13-NOV-1997; 97US-0065311P.
21-NOV-1997; 97US-0066120P.
24-NOV-1997; 97US-0066466P.
24-NOV-1997; 97US-0066772P.
11-DEC-1997; 97US-0069335P.
12-DEC-1997; 97US-0069425P.
17-DEC-1997; 97US-0069870P.
18-DEC-1997; 97US-0068017P.
10-MAR-1998; 98US-0077450P.
11-MAR-1998; 98US-0077632P.
11-MAR-1998; 98US-0077649P.
20-MAR-1998; 98US-0078866P.
20-MAR-1998; 98US-0078939P.
27-MAR-1998; 98US-0079664P.
27-MAR-1998; 98US-0079786P.
31-MAR-1998; 98US-0080107P.
31-MAR-1998; 98US-0080194P.
01-APR-1998; 98US-0080327P.
01-APR-1998; 98US-0080333P.
08-APR-1998; 98US-0081049P.
08-APR-1998; 98US-0081070P.
09-APR-1998; 98US-0081195P.
15-APR-1998; 98US-0081838P.
21-APR-1998; 98US-0082568P.
21-APR-1998; 98US-0082569P.
22-APR-1998; 98US-0082704P.
22-APR-1998; 98US-0082797P.
28-APR-1998; 98US-0083322P.
29-APR-1998; 98US-0083495P.
29-APR-1998; 98US-0083496P.
29-APR-1998; 98US-0083499P.
29-APR-1998; 98US-0083559P.
05-MAY-1998; 98US-0084366P.
06-MAY-1998; 98US-0084414P.
07-MAY-1998; 98US-0084639P.
07-MAY-1998; 98US-0084640P.
07-MAY-1998; 98US-0084643P.
15-MAY-1998; 98US-0085579P.
15-MAY-1998; 98US-0085580P.
15-MAY-1998; 98US-0085582P.
15-MAY-1998; 98US-0085700P.
18-MAY-1998; 98US-0086023P.

250 CT-TCCAGGC-----AGGCTCTATAAGTGACC-----GCC-- 278
294 CTGGCCATGCTCTCCCTTGGGCCCACTCAGTCAACCAAGACCCAGATTCTCCAGGGCCG 353
279 -GGGCGAGCGTGGCGGC-----GT-TGCAAGTCACTGTAAGGACTCTCTTTGGTT 328
354 GGTTCAACCTCAACACACACACAGAGTCTGCCATCCACAGGCGCTTCCAGCACCTGGTT 413
329 TTCTTTCTCTTTGG-----GGC-----ACCT-----CTGGAC-----TCACCTCCCCAGG 367
414 CACTCATGACTGTTCACCAAGACCTGACCTGAGATGGGAAGTGCCTCTTCGTC 473
368 ATGAAGCGGTGAGCCCGGTGC-----GGC-----TGCTACGAGCGGTGTGCT 413
474 AAGAAGAGAGTG-----CAGCTGCAGGCAAAATTTCTTGGGCAATCTCAAGAGCGCTGTATGA 530
414 GC-----CTGT-----CGGA-----ACGAGTCTGGCCATCCCGGGCGGCGAGG--AAG 457
531 GCAGAAGTCTTTCTACAGATTCTC-CAACCCCTCCATTGCCCAGG--CGAGGATCAAC 587
458 GGCCCGGACGTGAGGAGCGCTGAGCTTGTGACGACATGAACCACTGC-----TACTC 513
588 AGCCATGTGAAGAAAGAGACCAAGGGAAGGTTGTAGACATAATCCAAGGCGCTTGACTT 647
514 CCGCTGC-----GGGAAGTGT-----ACCCGGAGTC-----CC-----GAGAGGAC--T 553
648 CTGAGCGCCATGTTCTGTGTAATCACAATTTCTTTAAAGCCAAAGTGGGAGAGCCCTTT 707
554 CAGCTT-----AG-----CCAGGTGG-----AA 571
708 CACCTTGAATATACAGAAAGAACTTCCCATCTCTGTGGCGGAGGCTCAGTGTGCA 767
572 ATCC-----TACAG--CGCGT-----CATCG-----ACTAC 595
768 GTCCCCATGATGACCAAGAAAGACAGTTCGTCTTTGGGTGGATACAGAGCTGGAAGTGC 827
596 ATTCTGACCTGCAAGTAG--TCTGCGCGA--GCC--AGCC--CTTGGAGC--CCCT-- 643
828 TTT--GTGCTGAGATGATTAACAAGGAGATGCGGTGCGCTCTTTGTCCTCCCTAGC 884
644 -----GAT--GGCCCC-----CA--CTTCCCATCCAGACAGCGGA----- 675
885 AAGGGCAAGATGAGGCAACTGGAACAGGCTTGTGACCCAGAACTGATATAAGTGGAGC 944
676 -GCTCGCTCGGAAGT--T-----TCATCTCCACGACACAAAGAGGACTTTTC 721
945 CACTCACTCCAGAAAGGTGATAGAGGTGTTCATCCCC-----AGATTTT-C 991
722 CA---CTGACTCGGCGGTGCTGACACCTCCAGAACGCGAGGTGCTGGCGCCGTTCTGC 778
992 CATTTCTGCTCTCTAC-AATCTGGAACCATCTCCCGAAGATG--GGCATCCAAATGCG 1048
779 CTGGGAGC-----CCGG-----GAACCTCTCTGCGCG----- 806
1049 CTTTGACAAAATGCTGATTTTCTGGAATGCAAGAGAGACTC-CTGCAAGTTCTTA 1107
807 ----AAGCCGAGCGGACAG--GGATG-----GGCC----- 829
1108 AAGCAACCAAGCTGTGCTGATGTCTAGTGAAGAGGCACTGAGGCCACAGCAGCTA 1167
830 -----CGACTTC--GCC-----TG--CCCACTTGAATTCACCAATCCCTT 868
1168 CCACCAACAGTTCTATGTCGATCGATGCAAGGATGTCCTCTT-ACITCACTGTCTCTTC 1226
869 CTGAGACTAACTGCTGCT-----CA-----GGAGCGAAGGACTGTGAAC----- 912
1227 AATAGGACCT--TCTGATGATGATTACAAATAAGCCACAGAGGATTTCTTCTTAG 1284
913 ----TGTA-----CTGGAAGCGCAGAGTACTCTGCGC-----ACCAGCTGGCGGAC 958
1285 GGAAGTGGAAATCCCACTAAATCTAGTGGGAAATGCGCTGTAACTGATGG----- 1339
959 GTCACCTGCT-----CCACCCCAACCC-----CAAGTTCTTAAGG--TC 996

1340 --CACATTGCTAATGCAAGAAATAACAAACCAACATCCCTCTTCTTCTGAGGGTGC 1397
997 TTTTCAGAGC-CTGGAGGTG-----TGGAAGAGTGGCT-----GC-----TCTCC 1036
1398 ATTTGACCCAGTGGAGCTGGATTCCGCTGGCAGGATGCCACTTCCAAAGGCTCAATCACC 1457
1037 AAACCTAT-----GCCAAGGC-----GGC-----GGCAGAG--CTG 1064
1458 AAACCATCAACAGGAGCCCGCAGTCACAGCCCAACACCCAGTCAGTGCCTT 1517
1065 GTCTTCTGG-TCTCTTGG-AGAAAGTTT-----TGTTC-----CTGATTT--A 1107
1518 TTCCACAAATTTCTCCAGGTAAGTCTCATGGATGTTGCTGGGTTACCATATTTCCA 1577
1108 T-----GAACCTCTATAATAGAG-----TATATAGGTTTGTACTTTTACAGG 1152
1578 TTCTTGGGGCTCCAGGAATGGAATACGCCAACCCAGGTTAGGCACTCTATTGCAGA 1637
1153 -----AAGTGACTTCTGTAACTGATGATGATATATAAATTTTATAAAA 1200
1638 ATTACATAACACATTTCAATAAACTAAATATGAATTCAAAAA 1690

RESULT 13
ABU98260
ID ABU98260 standard; protein; 1743 AA.
XX
AC ABU98260;
XX
DT 30-JUL-2003 (first entry)
XX
DE Novel human secreted and transmembrane protein #3.
XX
KW Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy;
chondrocyte stimulator; tumour; adrenal tumour; lung tumour;
colon tumour; breast tumour; prostate tumour; rectal tumour;
cervical tumour; liver tumour; chromosome identification.
XX
OS Homo sapiens.
XX
FN US2003017544-A1.
XX
PD 23-JAN-2003.
XX
PF 21-JUN-2002; 2002US-00176915.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078866P.
PR 20-MAR-1998; 98US-0078939P.

PR	27-MAR-1998;	98US-0079664P.	PR	24-JUN-1998;	98US-0090540P.
PR	27-MAR-1998;	98US-0079786P.	PR	25-JUN-1998;	98US-0090676P.
PR	31-MAR-1998;	98US-0080107P.	PR	25-JUN-1998;	98US-0090678P.
PR	31-MAR-1998;	98US-0080194P.	PR	25-JUN-1998;	98US-0090688P.
PR	01-APR-1998;	98US-0080327P.	PR	25-JUN-1998;	98US-0090690P.
PR	01-APR-1998;	98US-0080333P.	PR	25-JUN-1998;	98US-0090694P.
PR	08-APR-1998;	98US-0081049P.	PR	25-JUN-1998;	98US-0090695P.
PR	08-APR-1998;	98US-0081070P.	PR	25-JUN-1998;	98US-0090696P.
PR	09-APR-1998;	98US-0081195P.	PR	26-JUN-1998;	98US-0090862P.
PR	15-APR-1998;	98US-0081838P.	PR	26-JUN-1998;	98US-0090863P.
PR	21-APR-1998;	98US-0082568P.	PR	26-JUN-1998;	98US-0091010P.
PR	21-APR-1998;	98US-0082569P.	PR	01-JUL-1998;	98US-0091359P.
PR	22-APR-1998;	98US-0082704P.	PR	01-JUL-1998;	98US-0091544P.
PR	22-APR-1998;	98US-0082797P.	PR	02-JUL-1998;	98US-0091478P.
PR	28-APR-1998;	98US-0083322P.	PR	02-JUL-1998;	98US-0091486P.
PR	29-APR-1998;	98US-0083495P.	PR	02-JUL-1998;	98US-0091626P.
PR	29-APR-1998;	98US-0083496P.	PR	02-JUL-1998;	98US-0091628P.
PR	29-APR-1998;	98US-0083499P.	PR	02-JUL-1998;	98US-0091632P.
PR	29-APR-1998;	98US-0083559P.	PR	04-JUL-1998;	98US-0094006P.
PR	05-MAY-1998;	98US-0084366P.	PR	04-AUG-1998;	98US-0092582P.
PR	06-MAY-1998;	98US-0084414P.	PR	10-AUG-1998;	98US-0095998P.
PR	07-MAY-1998;	98US-0084639P.	PR	10-AUG-1998;	98US-0096012P.
PR	07-MAY-1998;	98US-0084643P.	PR	17-AUG-1998;	98US-0096757P.
PR	15-MAY-1998;	98US-0085579P.	PR	17-AUG-1998;	98US-0096766P.
PR	15-MAY-1998;	98US-0085580P.	PR	17-AUG-1998;	98US-0096867P.
PR	15-MAY-1998;	98US-0085582P.	PR	17-AUG-1998;	98US-0096891P.
PR	15-MAY-1998;	98US-0085700P.	PR	17-AUG-1998;	98US-0096897P.
PR	18-MAY-1998;	98US-0086023P.	PR	18-AUG-1998;	98US-0096949P.
PR	22-MAY-1998;	98US-0086392P.	PR	18-AUG-1998;	98US-0097022P.
PR	22-MAY-1998;	98US-0086486P.	PR	26-AUG-1998;	98US-0097552P.
PR	28-MAY-1998;	98US-0087098P.	PR	26-AUG-1998;	98US-0097954P.
PR	28-MAY-1998;	98US-0087208P.	PR	26-AUG-1998;	98US-0097955P.
PR	02-JUN-1998;	98US-0087609P.	PR	26-AUG-1998;	98US-0097971P.
PR	02-JUN-1998;	98US-0087759P.	PR	26-AUG-1998;	98US-0097974P.
PR	03-JUN-1998;	98US-0087827P.	PR	26-AUG-1998;	98US-0098014P.
PR	04-JUN-1998;	98US-0088025P.	PR	01-SEP-1998;	98US-0098716P.
PR	04-JUN-1998;	98US-0088028P.	PR	01-SEP-1998;	98US-0098723P.
PR	04-JUN-1998;	98US-0088029P.	PR	02-SEP-1998;	98US-0098803P.
PR	04-JUN-1998;	98US-0088033P.	PR	02-SEP-1998;	98US-0098821P.
PR	04-JUN-1998;	98US-0088326P.	PR	02-SEP-1998;	98US-0098843P.
PR	05-JUN-1998;	98US-0088167P.	PR	09-SEP-1998;	98US-0098603P.
PR	05-JUN-1998;	98US-0088202P.	PR	09-SEP-1998;	98US-0098602P.
PR	05-JUN-1998;	98US-0088202P.	PR	10-SEP-1998;	98US-0098741P.
PR	05-JUN-1998;	98US-0088212P.	PR	10-SEP-1998;	98US-0098754P.
PR	05-JUN-1998;	98US-0088217P.	PR	10-SEP-1998;	98US-0098763P.
PR	09-JUN-1998;	98US-0088555P.	PR	10-SEP-1998;	98US-0098763P.
PR	10-JUN-1998;	98US-0088722P.	PR	10-SEP-1998;	98US-0098812P.
PR	10-JUN-1998;	98US-0088738P.	PR	15-SEP-1998;	98US-0100388P.
PR	10-JUN-1998;	98US-0088740P.	PR	16-SEP-1998;	98US-0100662P.
PR	10-JUN-1998;	98US-0088740P.	PR	16-SEP-1998;	98US-0100664P.
PR	10-JUN-1998;	98US-0088811P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;	98US-0088824P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;				

PR 30-SEP-1998; 98US-0102570P.
 PR 30-SEP-1998; 98US-0102571P.
 PR 01-OCT-1998; 98US-0102684P.
 PR 01-OCT-1998; 98US-0102687P.
 PR 02-OCT-1998; 98US-0102965P.
 PR 06-OCT-1998; 98US-0103258P.
 PR 06-OCT-1998; 98US-0103449P.

 Query Match 41.1%; Score 3101.5; DB 6; Length 1743;
 Best Local Similarity 41.8%; Pred. No. 2.1e-130;
 Matches 725; Conservative 0; Mismatches 421; Indels 587; Gaps 102;

 QY 1 GATCTGGG-----GTGCTGCCAGGAAAAGCAAAATCTTGAAGTTAATGTTTGGAG 52
 DB 12 GGACAGGGCAGGAGAGAGAAAGCACTGCAAGA-CATAATTTGTTCCAAATGGCATC--- 67
 QY 53 TGATTTTAA-----AATCCTTCTCGCGAGAGGCCCGCTCT---CCCGGTAT-CAG 102
 DB 68 TTACCTTTATGGAGTACTCTTCTCTGTTG-----GCTCTGTCTCCATCTACTG 118
 QY 103 CGTTCTCTCATCTTTTGAATCC---GGGCTCGCGGTCTTCGGGTC---AGACAGCGG 157
 DB 119 TGTGTCCCGGCAATGCCCCAGTGCATACCCCGCCCTCTCTCCACAAAGAGCACCC- 177
 QY 158 GAGGAAGCC-----TGTTTGAATTTAAGGGGCTGTG---AAGCCCGAGG---GCCGG 205
 DB 178 ---CTGCTTCACAGGTGATTCCTTCACACCGACTTTGCTTCGCTATACCGCAGG 233
 QY 206 CGGG-----GGCGGGCGGAGCGCGGCATTT-----TGAAATAAAGAGGCTGC 249
 DB 234 CTGGTTTGGAGACCCCGAGTCAGAAATCTTCTCTCCCTGTGAGTGTCTCCACTTCC 293
 QY 250 CT-TCCAGGC-----AGGCTCTAAGTGACC-----GCC-- 278
 DB 294 CTGCGCATGTCTCCCTTGGGGCCCACTCAGTCACCAAGACCCAGATTCACAGGCGCTG 353
 QY 279 -GGGGCGAGCGTGGCGC-----GT-TGCAAGTCACTGTAGCGGACTTCTTTTGGTT 328
 DB 354 GGCTTCAACTCACACACACAGAGTGTGCCATCCACAGGGCTTCCAGCACCTGGTT 413
 QY 329 TTCTTTCTTTGG-----GGC-----ACCT---CTGGAC-----TCATCCCCCAGC 367
 DB 414 CACTCACTGACTGTTCACGAAAGACCTGACCTTGAAGATGGGAAGTGCCTCTTCTGTC 473
 QY 368 ATGAAGCGCTGAGCGCGTGC---GC-----GGC---TGCTACGAGGGGCTGTGCT 413
 DB 474 AAGNAGAGCTG---CAGCTGCAGGCAAAATTTCTTGGCAATGTCAAGAGGCTGTATGAA 530
 QY 414 GC-----CTGT---CGGA-----ACGAGTCTGGCGATCGCCCGGGCGCGAGG--AAG 457
 DB 531 GCAGAGCTCTTTTCTACAGATTTCTC---CAACCCCTCCATTTGCCAGG---CGAGGATCAAC 587
 QY 458 GGCCCGGCAGCTGAGGAGCGCTGAGCTTCTGGAGCAGATGAACCACTGC-----TACTC 513
 DB 588 AGCATGTGAAGAAGAGACCCAAAGGAAGTTGTAGACATATCCAAAGGCTTGACCTT 647
 QY 514 CCGCTGTC---GGGAATCTGGT---ACCGGAGTC-----CC-----GAGAGGCAC--T 553
 DB 648 CTGACGCCCATGTTCTGTGTGAATCAATTTCTTTAAAGCCAAAGTGGGAGAGGCCCTTT 707
 QY 554 CAGCTT-----AG-----CCAGGTGG-----AA 571
 DB 708 CACTTGAATATACAGAAAGAACTTCCCATTTCTGTGGCGCAGCAGGTCACTGTGCA 767
 QY 572 ATCC-----TACAG---CGCCT-----CATCG-----ACTAC 595
 DB 768 GTCCCCATGATGCACCAAGAGACAGTCTGCTTTTGGGTGGATACAGAGCTGAACTGC 827
 QY 596 ATTCTGACCTGCGAGTAG---TCCTGGCGA-GCC---AGCC---CCTGGACC-CCCT--- 643
 DB 828 TTT---GTGCTGCAGATGGATTAACAGGGAGATGCGGTGGCCCTCTTTGCTCCCTAGC 884
 QY 644 -----GAT---GGCCCC-----CA---CCTTCCCATCCACAGACCGCA----- 675

Db 885 AAGGCAAGATGAGGCAACTGGAACAGGCTTGTCAAGCCAGAACACTGATAAAGTGAGC 944
 QY 676 -GCTCGCTCCGGAACCTTG-----TCACTCCCAACGACAAAGAGGCTTTTGC 721
 Db 945 CACTCACTCCAGAAAGGTGGATAGAGTGTTCATCCC-----AGATTTT-C 991
 QY 722 CA---CTGACTCGGCCCTGTCTGACACCTCCAGAACGACGAGGTGCTGGCGCCCTTCTGC 778
 Db 992 CATTTCGTCTCTCTAC-AACTGTGAACACCATCTCCCGAAGATG--GGCATCCAAAATGC 1048
 QY 779 CTGGGACC-----CCGG-----GAACCTCTCTCTGCGG----- 806
 Db 1049 CTTTGACAAAATGCTGATTTTCTGGAATGCAAGAGAGACTC-CTGCAGGTTTCTA 1107
 QY 807 ---AAGCGGACGGCAG---GGATG-----GGCC----- 829
 Db 1108 AAGCAACCCACAGAGCTGTGTGATGTCAAGAGGAGCAGTGAAGGCGACAGCAGCTA 1167
 QY 830 -----CCAACTTC---GCCC-----TG---CCCCTTGTACCAAAATCCCTT 868
 Db 1168 CCACACCAAGTTTCAATGTCGATCGAAGAGATGTTCTCTT-ACITCACTGTCTCTTC 1226
 QY 869 CTTGAGAGCTAAACCTGTGTCT-----CA---GGAGCGAAGGAGTGTGAAT----- 912
 Db 1227 AATAGGACCT--TCCTGATGATGATTACAAATAAGCCACAGCGGTATTCTCTTTCTAG 1284
 QY 913 ---TGTAG-----CCTGAAGCCAGAGTGTCTGGCC-----ACCAGCTGGGCGAC 958
 Db 1285 GGAAGTGGAAATCCCACTAAATCTAGTGGGAAATGGCCTGTTAACTGATGG----- 1339
 QY 959 GTCAACCTGTCT-----CCCAACCCCAACCC-----CAAGTTCTTAAG--TC 996
 Db 1340 ---CACATTTCTAATGACAGAAATAACAAACCATCTCTTCTCTGTTCTGAGGTGC 1397
 QY 997 TTTTCAGAGC-GTGGAGGTG-----TGAAGAGGTGGCT-----GC-----TCTCC 1036
 Db 1398 ATTTGACCCAGTGGAGCTGGAATTCGCTGGCAGGATGCCACTTCCAAGGCTCAATCACC 1457
 QY 1037 AAATCTAT-----GCCAAGGC-----GGC-----GGCAGAG--CTG 1064
 Db 1458 AAACATCAACAGGAGCCCGCAGTCACAGCCACACCCATTAACCCAGTCAGTGCCTT 1517
 QY 1065 GTCTTCTGG-TCTCCTTGG-AGAAAGTTC-----TGTTCG-----CCTGATTT--A 1107
 Db 1518 TTCCACAAATTTCTCCAGGTAACCTAGTCTCATGGGATGTTGCTGGGTACCATATTTCCA 1577
 QY 1108 T-----GAACTCTATAATAGAG-----TATATAGTTTGTACCTTTTTCAGG 1152
 Db 1578 TTCTTGGGCTCCAGGATGGAATACGCCAACCCAGGTTAGGACCTCTATTCAGA 1637
 QY 1153 ---AAGGTGACTTCTGTAACTGATGTATATTAACCTTTTATATAA 1200
 Db 1638 ATTACATAACACATTCATATAAACTAAATATGAATCAAAAAA 1690

RESULT 14
 ABU91966
 ID ABU91966 standard; protein; 1743 AA.
 XX
 AC ABU91966;
 XX
 DT 11-AUG-2003 (first entry)
 XX
 DE Novel human secreted and transmembrane PRO protein #3.
 XX
 KW Human; gene therapy; chromosome identification; tissue typing.
 XX
 OS Homo sapiens.
 XX
 PN US2003027277-A1.
 XX
 PD 06-FEB-2003.

XX	21-JUN-2002;	2002US-00176985.	PR	05-JUN-1998;	98US-0088217P.
XX	18-SEP-1997;	97US-0059263P.	PR	09-JUN-1998;	98US-0088655P.
PR	18-SEP-1997;	97US-0059266P.	PR	10-JUN-1998;	98US-0088722P.
PR	17-OCT-1997;	97US-0062250P.	PR	10-JUN-1998;	98US-0088738P.
PR	21-OCT-1997;	97US-0063486P.	PR	10-JUN-1998;	98US-0088740P.
PR	24-OCT-1997;	97US-0063120P.	PR	10-JUN-1998;	98US-0088811P.
PR	24-OCT-1997;	97US-0063120P.	PR	10-JUN-1998;	98US-0088824P.
PR	28-OCT-1997;	97US-0063121P.	PR	10-JUN-1998;	98US-0088825P.
PR	28-OCT-1997;	97US-0063540P.	PR	11-JUN-1998;	98US-0088826P.
PR	28-OCT-1997;	97US-0063541P.	PR	11-JUN-1998;	98US-0088861P.
PR	28-OCT-1997;	97US-0063544P.	PR	11-JUN-1998;	98US-0088863P.
PR	28-OCT-1997;	97US-0063564P.	PR	12-JUN-1998;	98US-0088876P.
PR	29-OCT-1997;	97US-0063734P.	PR	12-JUN-1998;	98US-0089105P.
PR	31-OCT-1997;	97US-0063870P.	PR	16-JUN-1998;	98US-0089512P.
PR	13-NOV-1997;	97US-0064103P.	PR	16-JUN-1998;	98US-0089512P.
PR	13-NOV-1997;	97US-0065311P.	PR	17-JUN-1998;	98US-0089514P.
PR	21-NOV-1997;	97US-0066120P.	PR	17-JUN-1998;	98US-0089538P.
PR	24-NOV-1997;	97US-0066468P.	PR	17-JUN-1998;	98US-0089598P.
PR	24-NOV-1997;	97US-0066772P.	PR	17-JUN-1998;	98US-0089653P.
PR	11-DEC-1997;	97US-0069335P.	PR	18-JUN-1998;	98US-0089908P.
PR	12-DEC-1997;	97US-0069425P.	PR	19-JUN-1998;	98US-0089952P.
PR	17-DEC-1997;	97US-0069870P.	PR	22-JUN-1998;	98US-0090246P.
PR	18-DEC-1997;	97US-0068017P.	PR	22-JUN-1998;	98US-0090252P.
PR	11-MAR-1998;	98US-0077450P.	PR	22-JUN-1998;	98US-0090254P.
PR	11-MAR-1998;	98US-0077632P.	PR	24-JUN-1998;	98US-0090429P.
PR	11-MAR-1998;	98US-0077643P.	PR	24-JUN-1998;	98US-0090435P.
PR	20-MAR-1998;	98US-0078886P.	PR	24-JUN-1998;	98US-0090444P.
PR	20-MAR-1998;	98US-0078939P.	PR	24-JUN-1998;	98US-0090461P.
PR	27-MAR-1998;	98US-0079664P.	PR	24-JUN-1998;	98US-0090535P.
PR	31-MAR-1998;	98US-0079786P.	PR	25-JUN-1998;	98US-0090540P.
PR	31-MAR-1998;	98US-0080107P.	PR	25-JUN-1998;	98US-0090676P.
PR	31-MAR-1998;	98US-0080194P.	PR	25-JUN-1998;	98US-0090678P.
PR	01-APR-1998;	98US-0080327P.	PR	25-JUN-1998;	98US-0090688P.
PR	01-APR-1998;	98US-0080333P.	PR	25-JUN-1998;	98US-0090690P.
PR	08-APR-1998;	98US-0081049P.	PR	25-JUN-1998;	98US-0090694P.
PR	08-APR-1998;	98US-0081070P.	PR	25-JUN-1998;	98US-0090695P.
PR	09-APR-1998;	98US-0081195P.	PR	26-JUN-1998;	98US-0090696P.
PR	15-APR-1998;	98US-0081838P.	PR	26-JUN-1998;	98US-00105413.
PR	21-APR-1998;	98US-0082568P.	PR	26-JUN-1998;	98US-0090862P.
PR	21-APR-1998;	98US-0082569P.	PR	26-JUN-1998;	98US-0090863P.
PR	22-APR-1998;	98US-0082704P.	PR	01-JUL-1998;	98US-0091010P.
PR	22-APR-1998;	98US-0082797P.	PR	01-JUL-1998;	98US-0091359P.
PR	28-APR-1998;	98US-0083322P.	PR	02-JUL-1998;	98US-0091544P.
PR	29-APR-1998;	98US-0083493P.	PR	02-JUL-1998;	98US-0091478P.
PR	29-APR-1998;	98US-0083496P.	PR	02-JUL-1998;	98US-0091486P.
PR	29-APR-1998;	98US-0083559P.	PR	02-JUL-1998;	98US-0091626P.
PR	05-MAY-1998;	98US-0084414P.	PR	02-JUL-1998;	98US-0091628P.
PR	07-MAY-1998;	98US-0084532P.	PR	04-AUG-1998;	98US-0094006P.
PR	07-MAY-1998;	98US-0084640P.	PR	10-AUG-1998;	98US-0095282P.
PR	07-MAY-1998;	98US-0084644P.	PR	10-AUG-1998;	98US-0095958P.
PR	15-MAY-1998;	98US-0085579P.	PR	17-AUG-1998;	98US-0096012P.
PR	15-MAY-1998;	98US-0085580P.	PR	17-AUG-1998;	98US-0096757P.
PR	15-MAY-1998;	98US-0085582P.	PR	17-AUG-1998;	98US-0096766P.
PR	15-MAY-1998;	98US-0085583P.	PR	17-AUG-1998;	98US-0096867P.
PR	18-MAY-1998;	98US-0086023P.	PR	17-AUG-1998;	98US-0096891P.
PR	22-MAY-1998;	98US-0086392P.	PR	18-AUG-1998;	98US-0096897P.
PR	22-MAY-1998;	98US-0086488P.	PR	18-AUG-1998;	98US-0096949P.
PR	28-MAY-1998;	98US-0087098P.	PR	18-AUG-1998;	98US-0096959P.
PR	28-MAY-1998;	98US-0087208P.	PR	18-AUG-1998;	98US-0097022P.
PR	02-JUN-1998;	98US-0087609P.	PR	26-AUG-1998;	98US-0097954P.
PR	03-JUN-1998;	98US-0087759P.	PR	26-AUG-1998;	98US-0097955P.
PR	04-JUN-1998;	98US-0088028P.	PR	26-AUG-1998;	98US-0097971P.
PR	04-JUN-1998;	98US-0088029P.	PR	26-AUG-1998;	98US-0097974P.
PR	04-JUN-1998;	98US-0088033P.	PR	01-SEP-1998;	98US-0098014P.
PR	04-JUN-1998;	98US-0088037P.	PR	01-SEP-1998;	98US-0098716P.
PR	05-JUN-1998;	98US-0088167P.	PR	02-SEP-1998;	98US-0098723P.
PR	05-JUN-1998;	98US-0088202P.	PR	02-SEP-1998;	98US-0098803P.
PR	05-JUN-1998;	98US-0088212P.	PR	02-SEP-1998;	98US-0098821P.
PR	05-JUN-1998;	98US-0088219P.	PR	09-SEP-1998;	98US-0098843P.
PR	05-JUN-1998;	98US-0088219P.	PR	10-SEP-1998;	98US-0099602P.
PR	05-JUN-1998;	98US-0088219P.	PR	10-SEP-1998;	98US-0099741P.


```

Db      1458 AAACCATCAACAGGGACCCAGTCACACGCCAACCCATTAAACCCAGTCAGTCCCTT 1517
QY      1065 GTCTTCTGG-TCTCCTTGG-AGAAAGGTTT-----TGTTC-----CCTGATT--A 1107
Db      1518 TTCCACAAATTCCTCCAGGTAACCTTTCATGGGATGTTGCTGGTTACCATATTGCCA 1577
QY      1108 T-----GAATCTATATATAGAG-----TATATAGTTTCTTACTTTTACAGG 1152
Db      1578 TTCTTGGGCTCCAGGAATGAATACGCCAGCCAGTTAGGCACCTCTATTGCGAGA 1637
QY      1153 -----AAGTGTACTTCTCTGTAACAATGCGATGTATATTAACCTTTTATAAAA 1200
Db      1638 ATTACATAACACATTCATAAACTAAATATGAATTCAAAAAATAAAAAA 1690

RESULT 15
ABU85270
ID ABU85270 standard; protein; 1743 AA.
XX
AC ABU85270;
XX
DT 30-JUN-2003 (first entry)
XX
DE Novel human secreted and transmembrane protein #3.
XX
KW Human; secreted and transmembrane protein; PRO; cytosolic; gene therapy;
KW chondrocyte stimulator; chromosome mapping; gene mapping;
KW transgenic animal; knock-out animal; tumour.
XX
OS Homo sapiens.
XX
XX US2003032114-A1.
XX
PD 13-FEB-2003.
XX
XX 20-JUN-2002; 2002US-00176919.
XX
FF 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059263P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 11-DEC-1997; 97US-0066772P.
PR 12-DEC-1997; 97US-0069335P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.

```

```

PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089633P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.

```


Db 992 CATTTCTGCTCTAC-AATCTGGAAACCATCTCCCGAAGATG--GGATCCAAATGC 1048
 Qy 779 CTGGAGC-----CCGG-----GAACCTCTCTGCGG----- 806
 Db 1049 CTTTGACAAATGCTGATTTTCTGGAATTGCAAGAGAGACTC-CCTGCAGTTTCTA 1107
 Qy 807 -----AAGCGGAGCGGAG-----GGATG-----GGCC----- 829
 Db 1108 AAGCAACCCACAAAGGCTGTGCTGATGTAGAGAGGGCACTGAGGCCACAGAGCTA 1167
 Qy 830 -----CCAACTTC--GCC-----TG--CCCACTTGACTTCACCAATCCCTT 868
 Db 1168 CCACCACCAAGTTTCATAGTCCGATCGAAGGATGGTCCCTCTT-ACTTCACTGTCTCTTC 1226
 Qy 869 CTGGAGACTAAACCTGGTCT-----CA---GGAGGAGGACTGTGAACT----- 912
 Db 1227 AATAGGACCT--TCCTGATGATTAACAATAAGCCACAGACGGTATTCCTTTCTAG 1284
 Qy 913 ----TGTAG-----CCTGAAGGCCAGAGCTAGCTCTGCCC-----ACCAGCTGGGCGAC 958
 Db 1285 GGAAGTGGAAATCCCACTAATCTAGTGGGAATGGCCTGTAACTGATG----- 1339
 Qy 959 GTCACCTGTCT-----CCACCCCAACCC-----CAAGTTCTAAG--TC 996
 Db 1340 --CACATTGTCTAATGCACAAAGAAATACAAACCAACATCCCTCTTTCTGTCTGAGGGTGC 1397
 Qy 997 TTTTCAGAGC-GTGGAGGTG-----TGAAGGAGTGGCT-----GC-----TCTCC 1036
 Db 1398 ATTTGACCCAGTGGAGCTGGATTCGCTGSCAGGGATGCCACTTCCAAGGCTCAATCACC 1457
 Qy 1037 AAACATAT-----GCCAAGGC-----GGC-----GGCAGAG--CTG 1064
 Db 1458 AAACCATCAACAGGGACCCAGTCACAAGGCCAACACCCATTAAACCCCACTCAGTGCCTT 1517
 Qy 1065 GTCTTCTGG-TCTCCTTGG-AGAAAGTTC-----TGTTGC-----CCTGATTT--A 1107
 Db 1518 TTCACAAATTCCTCCAGGTAACTAGCTTCATGGGATGTGCTGGGTACCATATTTCCA 1577
 Qy 1108 T-----GAACTCTATAATAGAG-----TATATAGGTTTGTACCTTTTACAGG 1152
 Db 1578 TTCCTTGGGCTCCAGGAATGGAATACGCCAACCCAGGTAGGCACTCTATTGCAGA 1637
 Qy 1153 -----AAGTGACTTCTGTAAACAATGCGATGTATATAACTTTTATAAA 1200
 Db 1638 ATTACAATAACACATTCAATAAACTAAATAATGAATTCAAAAAATAAAAA 1690

Search completed: May 7, 2004, 15:08:41
 Job time : 102 secs

May 11 10:50:01 1967

History

10

US-09-548-372D-13			
Query Match 41.4%; Score 3123.5; DB 4; Length 2088;			
Best Local Similarity 37.5%; Pred. No. 8.5e-138;			
Matches 780; Conservative 0; Mismatches 403; Indels 895; Gaps 126;			
Qy	1	GATCTGGGTGCTGCAGAGAAAGAAATTCGAACT	52
Db	21	GCTCTGCTGGCGCGCTGACCGCTCGGCGCTGGAGGTACCCACTGATGATGATGCTGG	80
Qy	53	--TGATTTTAAATCC--TTGC-----TGCGGAGAG-----GCCC--GCCTCT	90
Db	81	CCTGCTGGCTGAACCCAGATGGCCATGTCCTGTGGGAGCTGAACATGACATGAATGT	140
Qy	91	CC-----CCGGT--ATCAGG--CTTCTCTCATTTCTTTGAATCCGCG--	128
Db	141	CCAGAAATGGAGTGGGATTCAGATCCATCAGGACCAAAACCTGATTTGATACCAAGA	200
Qy	129	----CTCC--GCGTCTT-----CGGCGTCAGACAG--CGGAGGAAGCCTGTTTG--CA	174
Db	201	AGGCATCTCTGATATTCGAGAGTCTACCTGAATGCGAGTCAACCAATGTGTAGA	260
Qy	175	ATTAAAGCGGTGTGAACGCCAGCGCGCGGCGGCGGCGCG--AGCGGG-----	226
Db	261	AGCCAAACCAACAGTGAACCTCAGAACTGGTGCRAAGCGGCGCGCGAGCAGTGAAGAC	320
Qy	227	-----CCATTTC-----ATAAAGAGGC-----GTGCC--	250
Db	321	CCATCCCATTTGTGATCCCTACCGCTGCTTAGTTGGTGGATTTGTAAGTGCCT	380
Qy	251	-----TTCCAGGCGG-----CTCTATAAGTG-----	272
Db	381	TCTCGTCTCTGACAAAGTGCAATTTCTTACACAGGAGAGGTGGATTTTGGGAACTCA	440
Qy	273	-----ACCGCGC-----GCGAGC--GTGCG-----CGCGTTGCGAGT	304
Db	441	TCTTCACTGCGACACCGTCCGCAAGAGAGATGATGAGTGAAGAGTACCACTTGATGA	500
Qy	305	C-AC---TGAGCGGAC-----TTT-----TTTTGTTTT--	330
Db	501	CTACGGCATGTTGTCGCTCGCGAATGACAAAGTTCCGAGGGTAGAGTTTGTGTGTG	560
Qy	331	-----CTTTCT-----CTTTGGGCACT--CTG-----GACTC--	357
Db	561	CCCACTGGCTGAAGAAAGTGAATGGAATTTGCTGTGATGCGAGGAGATGATTCGGA	620
Qy	358	-----ACTC---CCAGCA-----TGAAAGC-----	375
Db	621	TGTCGTGGGGCGGACAGACAGACTATGAGATGGAGTGAAGACAAAGTAGTAGA	680
Qy	376	-----GCTGAG-----CCGCTG-----CCGCTG--	394
Db	681	AGTAGCAGAGAGGAAGTGTGAGGTGGAGGAAGAAAGACCGGATGATGACGAGGA	740
Qy	395	TGCT-----ACGAGCGGTG-----TGCTG-----CTGTGCTG--	421
Db	741	CGATGAGGATGCTGATGAGTAGAGGAGAGGCTGAGGACCTTACGAAGAGCCACAGA	800
Qy	422	--GAAC--GCAGTCTGGCCATCGCC-----CGGGCCGAGGGAAGG-----	459
Db	801	GAGAAACCAACAGCATTTGCCACCAACCAACCAACCAACCAACCAACCAACCAACCA	860
Qy	460	-----CC--CGGCGCTG--AGGAGCGGTGA--GC--TTG-----CTGAGCGAC	496
Db	861	TCGAGTTCTTACACAGAGCCAGTACCCCTGATGCGGTGCAAGATCTCGAGACACC	920
Qy	497	-----ATGAACCACTGTACTTCCCGCTGC-----GGGAATGTGTA	532
Db	921	TGGGATGAGAAATGAACATGCCATTTCCAGAAAGCCAAAGAGAGAGGCTTGAGCCAA	980
Qy	533	CCCGGA-----GTCCC-----GAGAGGCA-----	551
Db	981	CCGAGAGAGAAATGCTCCAGGTGATGAGAGAAATGGAAGAGCGAGCGTCAAGCAAGA	1040

Qy	552	-----CT-----CAGCTTAGCCAGGTG-----GAATCCT--	576
Db	1041	CTTGCTTAAGCTGTATAGAGAGCAATTCAGCACTTTCCAGGAAAGTGAATCTTT	1100
Qy	577	-----ACAGC-----GGTTCATCGACT-----ACATTTCTCGACCTGC--AGGT	612
Db	1101	GGACAGGAGACGCAACGAGAGACAGAGCTGGTGGAGACACATGSCCAGAGTGA	1160
Qy	613	AGTCCTG-----GCGG--AGCCAGCCCTTGA-----CCCC--CTGATGGC	649
Db	1161	AGCCATGCTCAATGACCGCGCGCTGGCCCTGGGAACTACATCACCGCTCTGAGGC	1220
Qy	650	--CCCACTTCCCATCCAGACAGCCGAGCTC-----GC-----TCCG-----G	686
Db	1221	TGTTCTCTCTCGGCTC--GTCA--CGTGTTCATATGCTTAAAGAGTATGTCGCGCAG	1276
Qy	687	AACCT-----TGTCATCT--CCACGACAAAGAGCTTTTTC-----	721
Db	1277	AACAGAGGACAGACACACACCTTAAAGCATTTTCGAGCATGTGCGCATGTGCGATCCCA	1336
Qy	722	-----CACTGACTCGG--CCGTGTCTCTG--ACACCTCC-----AGAACGCA	758
Db	1337	AGAAAGCGCTCAGATCCGCTCCAGGTATGACACACCTCCGTGTGATTTATGAGCGCA	1396
Qy	759	-GPTGCTGGCGCGCTTCTGC--CTGGGACCCCGGAACTT--CTCTGCGG-----GAAGCC	811
Db	1397	TGAATCAGTCT	1451
Qy	812	GGAGCGGAGGATGGG-----CCCACT-----TCGCCCT--GCCCA	847
Db	1452	GGATGAAGTTGATGAGTCTCTCAGAAAGAGCAAACTATTTCAGATGACCTCTTGCCAA	1511
Qy	848	CTTGACTT---CACCAA--ATCCCTTCTGGAGACTAAAC--CTGGTGC-----TCAG	893
Db	1512	CATGATTTAGTGAACCAAGGATCAGTTAC--GGAACGATGCTCTCATGCCATCTTTGACCG	1570
Qy	894	GAGCGAAG--ACTGTGAA-----CTT-----GT--AGCTGAA-----G	924
Db	1571	AAAGCAAAACACCGTGGAGCTCTCTCCGTGAATGAGAGTTTCAGCTGGACGATCTCC	1630
Qy	925	AGCC-----AGAGCTAGCTCTG--GCCA--CCAGCTGGG-----CGAGCT-----	960
Db	1631	AGCCGTGCTATCTTTTGGGGTGAATCTGTGCGAGCAACAGAAAGTGAAGTGGAGC	1690
Qy	961	-----CACCTGCT--CCACC-----CCACCC--CCAAAGTTCTAAG--GT	995
Db	1691	CTGTTGATGCGCGCTCTGCTGCGAGCGAGACTGACCTCGACAGGTTCTGGGTGA	1750
Qy	996	C--TTTTCAGAGC--GTGGAG-----GTGTGGAAGGA-----GTGGCT	1029
Db	1751	CAATATCAAGACGAGGAGATCTCTGAAGTGAAGATGGATGAGAAATTCGACATGACT	1810
Qy	1030	-----GCTC-----TCCAAA-----CTATGC-----	1045
Db	1811	CAGGATATGAAGTTTCATCAAAATTTGGTGTCTTTGAGAGATGTGGGTCAAACA	1870
Qy	1046	-----CAAGGC--GGCGG-----CAGAGC-----TGCTCTCTGCT	1074
Db	1871	AAGTGCATATGAGTCTGATGCTGGCGGTGTGTCATAGCGACAGTATCTTC--AT	1928
Qy	1075	CTCTCTGAGA-----AAGTTCTGTCCTGCTGATTTATGAACTCTATATAGATA	1126
Db	1929	CACCTTGTGATGCTGAAGAGAAACAG--TACAC-----ATCCATTC--ATCATGTTGTG	1980
Qy	1127	TATAGTTTCTACTCTTTTTC-----AGGAAGTGAATTT--TC-----TGTAACAATGC	1175
Db	1981	GTGAGGTTGACGCGCTGTGTCACCCAGAGAGGCGCCCTGTCCAGATGCAGCAGAAC	2040
Qy	1176	G---ATGATATTAAC-----TTTTTATAAAGTT	1203
Db	2041	GGCTACGAAATCCAACTACAGTCTCTTGAGCAGAT	2078

RESULT 4
US-09-548-367D-13
; Sequence 13, Application US/09548367D
; Patent No. 6440698
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 29915/6280H
; CURRENT APPLICATION NUMBER: US/09/548,367D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 2088
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-367D-13

621	TG	TCTGGTGGGGGAGCAGACACAGACTATGCAGATGGGAGTGAAGCAAAAGTAGTAGA	680
376	QY	-----GCTGAG-----CCGGTG-----CGCGGC	394
681	DB	AGTAGCAGAGGAGGAAGTGCCTGAGTGGAAAGAAAGACCGATGATGACAGGA	740
395	QY	-----ACGAGCGGTG-----TGCTG-----CCTGTGCG-----	421
741	DB	CGATGAGGATGGTGATGAGTATAGGAGAGAGGCTGAGGAACTCTACGAGAGCCACAGA	800
422	QY	-----GAAC-----GCAGTCTGGCCATCGCC-----CGGGCCGAGGGGAAGGG-----	459
801	DB	GAGAACCCACGACATTCGCCACACACACACACACACACAGAGTCTGTGGAGAGCGTGT	860
460	QY	-----CC-----CGGACGCTG-AGGAGCGCTCA-GC--TTG-----CTGACGAC	496
861	DB	TCGAGTTCCTTACAACAGCAGCAGTACCCCTGATGCCGTTGACAAGATATCTCGAGAC	920
497	QY	-----ATGAACCACTGCTACTCCGCGCTGC-----GGAACTGTGTA	532
921	DB	TGGGATGAGATGAACATGTCCTATTTCCAGAAAGCCAAAGAGAGGCTTGAGCCAA	980
533	QY	-----GTCCC-----GAGAGGCA-----	551
981	DB	CCGAGAGAGATGTCGCCAGGTCAATGAGAGAAATGGGAGAGGCGAGACGTCAAGCA	1040
552	QY	-----CT-----CAGCTTAGCCAGGTG-----GAAATCTCT-	576
1041	DB	CTTGCTTAAGCTGATAAGAGGCAGTATCCAGCATTTCCAGGAGAAAGTGGATCTTT	1100
577	QY	-----ACAGC-----GGTCACTGACT-----ACATCTCGACCTGC-AGGT	612
1101	DB	GGAACAGGAGCAGCCACGAGAGACAGCAGCTGTGTGGAGACACATGCCCAGAGTGA	1160
613	QY	AGTCCCTG-----GCCG-AGCCAGCGCCCTGGA-----CCCC--CTGATGCG	649
1161	DB	AGCATGCTCAATGACCGCGCGCCCTGCGCTTGAGAACTACATCACCGCTCTGCAGC	1220
650	QY	-----CCCCACCTTCCCATCCAGACAGCGAGTC-----GC-----TCCG-----G	686
1221	DB	TGTTCTCTCTCGGCGCTC--GTCA--CGTGTCAATATGCTAAAGAGTATGTCCGCGC	1276
687	QY	AACT-----TGTCATCT--CCAAACGACAAAGGAGCTTTTGC-----	721
1277	DB	AACAGAGGACAGACAGCACACCTTAAGCATTTCCAGCATGTGCGCATGGTGGATCCCA	1336
722	QY	-----CACTGACTCG--CGGTGCTCTG--ACACCTCC-----AGAACGCA	758
1337	DB	AGAAAGCGCTCAGATCCGCTCCAGGTTATGACACACCTCGTGTGATTATGAGCGCA	1396
759	QY	GGTGCTGGCGCCCGTCTGCTG-CTGGAGACCCCGGAACTCTCTGCTGCG--GAAAGC	811
1397	DB	TGAATCAGTCTCTCTCCCTGCTCT--ACAACTG--CCTGCAGTGCAGGAGATTCA	1451
812	QY	GGACGGCAGGATGGGC-----CCCACT-----TGCCCT-CCCCA	847
1452	DB	CGATGAATGATGAGCTGCTTCAGAAAGACAAACTATTCAGATGACGTCTTGCCCAA	1511
848	QY	CTTGACTT--CACCA--ATCCCTTCTCGAGACTAAAC-CTGCTGC-----TCAG	893
1512	DB	CATGATTAGTGAACCAAGATCAGTTAC--GGAACGATGCTCTCATGCCATCTTGACCG	1570
894	QY	GAGCGAAGG--ACTGTGAA--CTT-----GT-----AGCTGAA-----G	924
1571	DB	AAACGAAACCACCGTGGAGCTCTCTCCCGTGAATGGAGAGTTTCAGCTTGACGATCTCC	1630
925	QY	AGCC-----AGAGCTAGCTCTG-GCCA-CCAGCTGG--CGAGCT-----	960
1631	DB	AGCGGTGCATCTTTTGGGCTGACTCTGTGCCGCAACACAGAAACGAGTTGAGC	1690
961	QY	-----CACCCTGCT-CCACCC-----CCACCC--CCAAGTTCTAAG--GT	995

Qy	722	-----CACTGACTCG--CCGTGTCTG--ACACTCC-----AGAACGCA 758
Db	1337	AGAAAGCCGCTCAGATCCCGTCCACAGTTATGACACACCTTCGTGTGATTTATGAGCGCA 1396
Qy	759	-GGTGTGGCGCCGCTTCTGC-CTGGAGACCCGGGAACCT-CTCCTGCGG-----GAAGCC 811
Db	1397	TGAATCAGTCTCTCTCCCTGCTCT--ACAACTG--CTTGCAGTGCCGAGAGATTCA 1451
Qy	812	GGAGGGCAGGGATGGG-----CCCAACT-----TCGCCCT-GCCCA 847
Db	1452	GSATGAAGTTGATGAGCTGCTTCAGAAAGAGCAAACTATTTCAGATGACGTCTTGCCCAA 1511
Qy	848	CTTGACTT-----CACCAA--ATCCCTTCTGGAGACTAAAC-CTGGTGC-----TCAG 893
Db	1512	CATGATTAGTGAACCAAGGATCAGTTAC-GGAAACGATGCTCTCATGCCATCTTTGACCG 1570
Qy	894	GAGCGAAG--ACTGTGAA---CTT-----GT--AGCCTGAA-----G 924
Db	1571	AAACGAAAACCAACCGTGGAGCTCTTCCCGTGAATGGAGAGTTTCAGCCCTGGAGACTCTCC 1630
Qy	925	AGCC-----AGAGCTAGTCTG-GCCA-CCAGCTGG-----CGAGCT-----960
Db	1631	AGCCGTGGCATTTCTTTTGGGCTGACTCTGTGCCAGCCAAACAGAAAACGAAGTTGAGC 1690
Qy	961	-----CACCTGCT-CCGACC-----CCACCC--CCAAGTTCCTAAG--GT 995
Db	1691	CTGTGTGATCCCGCCCTGCTGCCAGCCGAGACTGACCACCTCGACCAGGTTCTCGGGTTGA 1750
Qy	996	C--TTTTTCAGAGC-GTGGAG-----GTGTGGAAGA-----GTGGGT 1029
Db	1751	CAAAATACAAGACGGAGGAGATCTCTGAAGTGAAGATGGATGCAGAATCCGACATGACT 1810
Qy	1030	-----GCTC-----TCCAAA-----CTATGC-----1045
Db	1811	CAGGATATGAAGTTTCATCATCAAAAATGGTGTCTTTTCAGAGAAGATGGGGTTCAACA 1870
Qy	1046	-----CAAGGC-GGCCGG-----CAGAGC-----TGGTCTTCTGT 1074
Db	1871	AAGGTCAATCATTTGGACTCATGGTGGCGGTGTGTGTCTATAGCACAGTATCTTC--AT 1928
Qy	1075	CTCCTTGGAGA-----AAGGTTCTGTGGCCCTGATTTATGACTCTATATAGATTA 1126
Db	1929	CACCTTGTGATGCTGAAGAAGAACAG-TACAC-----ATCCATTC-ATCATGTGTG 1980
Qy	1127	TATAGGTTTGTACTTTTTTATC-----AGGAAGGTGACTT-TC-----TGTAAACATGC 1175
Db	1981	GTGGAGGTTGACGCGCTGTCACCCAGAGGAGCGCCACCTGTCCAAGATGCAGCAGAAC 2040
Qy	1176	G---ATGTATATTAAAC-----TTTTTATAAAGTT 1203
Db	2041	GGCTACGAAATCCAACTACCACTCAAGTCTTTTGAGCAGAT 2078

RESULT 6

```

US-09-627-650B-9
/ Sequence 9, Application US/09627650B
/ Patent No. 6406872
/ GENERAL INFORMATION:
/ APPLICANT: Bamber, Bruce
/ APPLICANT: Jorgensen, Erik
/ TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
/ TITLE OF INVENTION: Methods Related Thereto
/ FILE REFERENCE: 21101.0009U3
/ CURRENT APPLICATION NUMBER: US/09/627,650B
/ CURRENT FILING DATE: 2000-07-28
/ PRIOR APPLICATION NUMBER: 09/436,063
/ PRIOR FILING DATE: 1999-11-08
/ PRIOR APPLICATION NUMBER: 60/107,727
/ PRIOR FILING DATE: 1998-11-09
/ NUMBER OF SEQ ID NOS: 50
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 9
/ LENGTH: 2601

```

```

; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-9

Query Match 41.2%; Score 3105; DB 4; Length 2601;
Best Local Similarity 35.9%; Pred. No. 7.2e-137;
Matches 801; Conservative 0; Mismatches 388; Indels 1040; Gaps 132;

QY 1 GATCTGGGGTGC-----TGC-CA-----GGAAGAAGCAAAATTCGGA 37
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 GATATGGACTTCACATTAGACTTCTACATGCGTCAAACTGGCAAGACCTCGACTAGCC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 38 GTTAATAGTTTTGA---GTGATTTTT---AAATC-----CTTGCTGGCGAGAGG-- 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 TTGCGAAGCTTGATTGGGACTTCCAAAGAAATCGACTCACTTACCGTCGGAGTAGAC 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 82 -----CCGGC-----TCTCCCC----- 94
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 TACTCGATAGACTGTGGAACCCGACACGTTCTTCCCAATGAAAGAAATCATTTCTTC 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 95 -----GGTATCAGCGC-TTCTCTCAIT-CTTTGAATC----- 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 CACTTGGCAACACACATAACTCGTTCCTTCTGTAICGAGGTGATGACGGTTTTATCT 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 ---CGCGGTCCGCGGT---CTTCGGCGT---CA---GACCAG-----CCGAGGAA 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 604 AGTCAAAGATTAACTCACTGCAACGTCCTCAATGCACTGTAAGCTGTTCGCAATGGA- 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 164 GCCTGTTTCAATTTAAGC-GGCTCTGAACGC-CCAGGGC--GGCGGGGC----- 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 --CTCTAACACTGTAACTGAAATTTAAAGTACGGTACGACGAAAGATATCGAC 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 213 -----GGGCGGAGGCGGCCAATTTTGAATAAGAGC-----GTGCTTTC 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TACTATTGGGGAAGAAGCGGACTGATTTGGAGATAACGGCTGTCAAGTTTGATACCTTC 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 254 CAG-----CGAG-----GCTCTA-TAAGTGA-----CCG---CCGGC 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 CAGTTCGCGAGTTTCAGCCACGCTGTATTTGTGATACAACTAAAGCCGAGACTCA 840
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 282 GCAG-----CGTGGCC--GGTTTGCAGGT-----C 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841 TCAGGAAATACGTACGCTCGCGCTGGAAGTAAATTGGTTCGAAATATGGGCTTCTAC 900
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 306 ACTGT-AGCG--GAC TTC-----TTTTGGTTTTCTTT----- 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 901 ACTATGAACATCGTCATCCCATCCATCCGATCGTCACCATATCTTGGGTATCATTTGG 960
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 335 -----CTC-----TTTGGG-----GCAC----- 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 961 TTGAATCGAAGACTTCGCGGCTCGAGTTGATTTGGGTGTGACTACTGTGTCAATG 1020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 348 -----CTCTGGACTC-ACTCCCCA-----GCATGAAGCGCT-----GAGCCCG 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1021 ACAACTCTGATCACTACAAACCAATAATTCGATGCCAAAAGTGTCTTATGCAAGGCTCTG 1080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 386 G-TGCCCGGC-----TGCTACGAGGGGTGTGCTGCT-----GTGCG-GAACGC--- 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1081 GATGTGTTCTTAATTTTTTGTTCGTAATGGTAATTC-GCCTGTTGCTCGAGTACGCCAT 1139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 428 -----AGTGTGCCATCGCCCGG---GCCAGGGAAGG--GCC-- 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1140 AGTATCTCATGATAAAGCACTGTGCTCCGACGGGAAAAACGAAGAAAGCCGCCGA 1199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 462 -----CGGCAGC-----TGAGGAGCCGCT-----GAGCTTCTGGACGACATGA 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1200 ACAACAGACGGAACGAGATGCCAATGTTCAACGGGAGCCCGAAGCCGCCCAATTA 1259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 -----CCACT-----GCTAC--TC-CGCGCTGGGGAA-CTG-GTACCCGG 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1260 TTCAACGAATGACACTTATGTGCAAAATTCGACGCTGCCAAAAGCTATGTACAGGC 1319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 538 AGTC-----CCGAGGGCAC---TCAGCTTAG-----CCA--GGTGAAATCC----- 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 1320 TGACTTGTACTTTGCGG---GACACAAATTCCTCTATGAATCCATTTGATGAGATCCGAGA 1376
Qy 576 -----TACAGCGGT---CATGA---CTACATTTCTCGACCTG---CAGGTAGTCC 617
Db 1377 AAATTGTGATGCGCGAGATCCAAATGATGCAACATCCAGCTTTGTACAGAGCGGC 1436
Qy 618 -----TGGCGG---AGC-CAGCC-----CTTGGAC-----CCC 641
Db 1437 ACATACGCTATGCGCGGTCCATTTGCGCGCGCGGAAAGGCTTTCCAAAGACATGCTGCCA 1496
Qy 642 CTGATGGCCCC-----ACC-----TTCCCA----- 662
Db 1497 AGATGGAGCGCTGCAAAATCGATAAGCTTAGCGGATACGGTTTCCATGTTCTTCTC 1556
Qy 663 -----TCCAGACAGCC-----GAGTCTCGCT---CCGGAACCTTGTG---A 695
Db 1557 TATCTTCAATAGTCTACTGTTGTATATGAATAATATCAAGCTTAACTCTGCGGACAA 1616
Qy 696 TCTCC-----AACGACAAAGG-AGCTTTTCCCACTGA-----C-TGGGCG--- 734
Db 1617 GATCCAGGAGACGACAGTGGCAGCAT-CCACTGATGCGATTGACGGCGGGAATC 1675
Qy 735 -----GTGTCC-----TGACACC---TCCAGAACGGAG--- 759
Db 1676 GAGTACAAATGCTGATCGTGAAGGAGCGGAATTTGTCAGACGCGTCAAGGCGGACGCG 1735
Qy 760 -----GTGCTGGCG-----CCC---GTCTGCTG---GGACCC----- 787
Db 1736 AACATCGAATCGGATTAAATTTCACTAAATCTGCCAAACGACACATGCGCAGC 1795
Qy 788 -----CGGGAAC---CTCTC---CTGC-----CGGAAG--- 809
Db 1796 ACTTCATCGGACCTACTCTGCTACGGTGTAGTTTCAATTTGATCGGACACGCGG 1855
Qy 810 -----CCGAGCGG-----CAGGGAT----- 824
Db 1856 TTCTACTTTCTCAATATTTTCCCTGCGACGCTGCTGTAGTTTATCATGATCTCA 1915
Qy 825 -----GGGCCC---CAACTTCG---CCC-----TGCCC 846
Db 1916 TTCTGGATCAATCGTGAATCGGCGCTTTCGCGAACCTTAATCGGTACGATGACGGTGCTC 1975
Qy 847 ACT-----TGAC-----TTACCA----- 860
Db 1976 ACTGAGACTATCTTATGCGGACCAATCGACGCTTCCACCAGTTGCTATGTAAA 2035
Qy 861 -----AATCC-----CTTC-----CTGG-----AGACTAA 880
Db 2036 GCGGTGATGATCTCTCGTTTCTGCTATCTTCTGTTATCTGCGTTGATCGAGTAC 2095
Qy 881 ACTCG-----TGCTCAGGACGGAAGCACTGTGAATTT---GTAGCCTGAAG-AGCCA 929
Db 2096 GCTGTGTTGCTACTCAAAAAGAGAACGAGGATCGTCGAGAGAGAGAGAGAGC 2155
Qy 930 GAGCT-----AGCTCTG---GCCACCAG-----CTGGCGGACGTCACCT--- 966
Db 2156 GAGCATAACTGCTCCCGGACACTGATATTTCTTACGACGTCGCGCTTGGCGAATGC 2215
Qy 967 -----GCTCCACC-----CCACCCCAAGTTCTAAGTCTTTTCA 1004
Db 2216 ACATGCAACGCGGCTCCAACTCGATCATCGCGCTCATCAAGCAGTGCATCTGTT 2275
Qy 1005 G-----CGTGGAGTGTGGAAGAGTGC-----TGCTC----- 1033
Db 2276 GTCAGTCAGTCATTTGACAT-CGTACGCGGTGCGCGGTTTCTTCTTTTCTATCTT 2334
Qy 1034 ---TCCAAACTATGCCAAGCGG---CGGACAGCTGTTCTTCTGTCT-CC-----TTGG 1082
Db 2335 GTTCAACATCTCTTCTGCTGANTTCTACTGTACAAATCCAAAGGCTGCGGTATATTAG 2394
Qy 1083 AGAA---AGGTT---CTGTTC-----CCGTA---TTTATGA-CTCTAT----- 1117

Db 2395 TGAACACAGGGTGACCGTTGCGATGCTCCAGACCTTCATTATCTCAATCCAATCTCT 2454
Qy 1118 -----AATAGATATAGT---TTGTAC-----CTTTTTTAC-----AGGA 1153
Db 2455 CATCATTTCCATTTGAAATATCTTTTCTTGACAGAGCGCTTTTCTGTTTTTTT 2514
Qy 1154 AGTGACTTTCTGTAC-----AATGC-----GATG-----TATATTAAAC 1189
Db 2515 TATTGATTATTTTACGGATTTTATAGATAATGACAGATGCTCATCTCAATAAAT 2574
Qy 1190 TTTTATAA 1198
Db 2575 TTATTTAA 2583

RESULT 7
US-09-436-063C-9
; Sequence 9, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-9

Query Match 41.2%; Score 3105; DB 4; Length 2601;
Best Local Similarity 35.9%; Pred. No. 7,2e-137;
Matches 801; Conservative 0; Mismatches 388; Indels 1040; Gaps 132;

Qy 1 GATCTGGGGTGC-----TGC-CA-----GGAAAAAGCAAAATCTGGAA 37
Db 364 GATATGGACTTCACATTAGACTTCTACATGCGTCAACGTCGCAAGACCTCGACTAGC 423
Qy 38 GTTAATGTTTGA---GTGATTTT---AAATC-----CTTGCTGGCGAGG-- 81
Db 424 TTGGAAAGTTGATTTGGGACTTTCCAAAGAAATGACTCACTTACCCTCGAGTAGAC 483
Qy 82 -----CCGCGC-----TCTCCCC----- 94
Db 484 TACCTGATAGACTGTGGAACCCGACACGTTCTTCCCAATCAAAAGAAATCATTTCTC 543
Qy 95 -----GCTATCAGCGC-TTCTCTCAT-CTTTGAATC----- 123
Db 544 CACTTGGCAACACACATAACTCGTTCTTCTGATCGGGTGTGGAACGTTTATCT 603
Qy 124 ---CGCGGCTCCGGGT---CTTCGGGT---CA---GACCAG-----CCGAGGAA 163
Db 604 AGTCAAAGATTAAACAGTCACTGCAACGTCGCAATGACCTGAAGCTGTTCCCAATGA- 662
Qy 164 GCGTTTTCAAATTTAAGC-GGGCTCTGAACGC-CCAGGGCC--GGCGGGGC----- 212
Db 663 ---CTCTCAACACTGTAACTGGAATTTGAAGCTACGGGTACGAGACGAAAGATATCGAC 720
Qy 213 -----GGGCGCGAGGGCGGCCATTTTGAATAAGAGCC-----GTGCTTTC 253
Db 721 TACTATTGGGGGAAGAAGCGGACTGATTTGGAGATAACGGCTGTCAAGTTTGATACCTTC 780
Qy 254 CAG-----GCAG-----GCTCTA-TAAGTGA-----CCG---CCGCG 281
Db 781 CAGTTGCGCGAGTTTCAGCCACGCTGTATTITGATACAACTAAAGCCGAGACCTCA 840

282	GCAG-	-----CGTGGC-	---CGGTTGCAGGT-	-----C	305	
841	TCAG	AAAAATAGTACGCTGCGCTGGAAGTAA	TATTGGTTCGAAATATGGCTTCTAC	900		
306	ACTGT-AGCG-	-GACTTC-	-----TTTGGTTTCTTT-	---334		
901	ACTATGAACATCGTCATCCATCCATCCATGATCGTACCA	CCATATCTTTGGGTATCA	TTTGGT	960		
335	-----CTC-	---TTTGGG-	-----GCAC-	---347		
961	TTGAATCGAAGACTTCGCGGCTCGAGTTGGATTGGGTG	TGACTACTGTCTCAATG	1020			
348	-----CTCTGGATC-ACTCCCA-	-----GCATGAAGCGCT-	-----GAGCCCG	385		
1021	ACAACTCTGATCACTACACCAATATTCGATGCCAAAAGT	GTCTTATGTCGAAGGTCTG	1080			
386	G-TGCGCGG-	-----TGCTACGAGCGGTGTGCTGCT-	-----GTGCG-GAAAGC-	---427		
1081	GATGTGTTCTTAAATTTTGTTCGTAATGGTATTC-	GCTCGTGTCTCGAGTACCCAT	1139			
428	-----AGTCTGGCCATCGCCCGG-	---GCCGAGGAAGG-	---GCC-	---461		
1140	AGTATCCTACATGAATAAACGACTGTGCTCCAGCGG	AAAAACGAAAGAAACGCCGCCGA	1199			
462	-----CGGACG-	---TGAGGACCGCT-	-----GAGCTTCTCGACACATGAA-	---501		
1200	ACAACAGCAGCAACGAGATGCCAATGTTCAACGCGAG	CGCCGGAAGCGCCCAATATAA	1259			
502	-----CCACT-	-----GCTAC-	---TC-CGCGCTGCGGAA-CTG-GTACCCGG	537		
1260	TTCATACGAAATGACACTTATGTGCGAAAATTCGAG	CGCTGCCAAAAGCTATGTACAGC	1319			
538	AGTC-	-----CCGAGAGGCAC-	---TCAGCTTAG-	---CCA-GGTGAAATCC-	---575	
1320	TGACTTGTACTTTGCGG-	---GACACAATTCCTCTATGAATCCATTTGAT	GCAGATCCAGA	1376		
576	-----TACAGCGCGT-	---CATCGA-	---CTACATCTCGACCTG-	---CAGGTAGTCC	617	
1377	AAATTGTGTTGCGGACGATTCCAATGATGCAACAT	CCACGTCTTGTACAGACGCGC	1436			
618	-----TGCGCG-	---AGC-GAGCC-	-----CCTGGAC-	---CCC	641	
1437	ACATAGCTATGCGCGCTCCATTCGCGCGCGCGA	AAAAAGCTTCCAGACATGCTGCCA	1496			
642	CTGATGCCCCC-	-----ACC-	-----TTCCCA-	---662		
1497	ACGATGACGCTGCRAAATCGATAAGTTAGCCGATAG	CGGTTCCTCATTTCTCTC	1556			
663	-----TCCAGACAGCC-	---GAGCTCGCT-	---CCGGAACCTTGTC-	---A	695	
1557	TATCTTCAATATAGTCTACTGGTTGTATGAANA	TCTAAGCTTAACCTCTCGHCAA	1616			
696	TCTCC-	-----AACGACAAAAAG-AGCTTTTGCCACTGA-	-----C-TCGGCC-	---734		
1617	GATCCAGGAGACACAGTGGCAGACAGAT-CCA	CTGATGCGTATTCGAGCGCGCAATC	1675			
735	-----GTGTCC-	-----TGACACC-	---TCCAGAACGAG-	---759		
1676	GAGTACAAATGGTTACGTGGAAGGACCGCAAT	TGTCGACAGCGGTCAAGCGCGACGCG	1735			
760	-----GTGCTGGGG-	-----CCC-	---GTTCTGCTGTG-	---GGACCC-	---787	
1736	AACATCGAATGTGAGTTATAAATTCACATAAAT	CTGCCAAAACGACACTTGCACGC	1795			
788	-----CGGGAAC-	---CTCTC-CTGC-	-----CGGAAG-	---809		
1796	ACTTCATCGGGAGCCTACTCTCGTCTACGGGTAG	TTTCATATTTGATTCGCGACAGCGGC	1855			
810	-----CCGACGCG-	-----CAGGGAT-	---824			
1856	TTCTACTTCTTCAAAATATTTTCCCTGCCAGCCT	CGTGTAGTTTATCATGATCTCA	1915			
825	-----GGGGCC-	---CRACTTGC-	-----TGCCC	846		

Db	1916	TTCTGGATCAATCGTGACTCGGGGCTTCGCGAA	CCCTAATCGGTACGATGACGGTGTCT	1975
Qy	847	ACTT-----TGAC-----TTACCA-----		860
Db	1976	ACTGAGACTCATCTTATGACCGGAACCAATCGACGCTTCCACCAAGTGCCTATGTAAAA		2035
Qy	861	-----AATCC-----CTTC-----CTGG-----AGACTAA		880
Db	2036	GCGGTGATGATATCTCTCGGTTCTGCTATCTCTGGTTATACTGCGGTTGATCGAGTAC		2095
Qy	881	ACCTGG-----TGCTGAGGAGCGAAGGACTGTGAACCTT-----GTAGCTGGAAG-AGCCA		929
Db	2096	GCCTGTGTTGCCCTACTCAAAAAAAGAAACGAGGATCGTCGGAAGAGAGAGAAGAACG		2155
Qy	930	GAGCT-----AGCTCTG--GCCACGAG-----CTGGGCGAGCTACACCT-----		966
Db	2156	GAGCAATAACTGCTCGCGACACCTGATATCTTTCACGAGCTCGCGCTTCCGGAATGC		2215
Qy	967	-----GCTCCACG-----CAACCCCAAGTTCTTAAGTCTTTTCAGA		1004
Db	2216	ACATGCAACGGGGTCCAACTCGGATCATCGCGCTCATCAAGCAGTCSAATCGAATCTGT		2275
Qy	1005	G-----CGTGGAGGTGTGAAGAGGTGGC-----TGCTC-----		1033
Db	2276	GTAGTCACAGTCACATGTACAT--CGTCAGCGTGGCGGTTTCTCTTGTTTTCATCTT		2334
Qy	1034	--TCCAAACTATGCCAAGCGG--CGGAGAGCTGGTCTCTTGGTCT--CC-----TTGG		1082
Db	2335	GTTCAACACTCTCTTCTGGTGATTTACTGTACAAATCCAAAGCGTCTCGGTATATTAG		2394
Qy	1083	AGAA-----AGTT--CTGTTC-----CCTGA--TTTATGAA-CTCTAT-----		1117
Db	2395	TGAACAGGGGTGACCGTTTGGATGCTCCAGACCTTCAATTAATCTCAATCCAACTTCCT		2454
Qy	1118	-----AATAGATATATAGTT--TTGTAC-----CTTTTTTAC-----AGGA		1153
Db	2455	CATCATTTTCCANTTCGAATATCTTTTCTTGCACAGAAGCGTTTTTCGTTTTTTTTT		2514
Qy	1154	AGGTGACTTCTGTAAC-----AATGC--GATG-----TATATTAAAC		1189
Db	2515	TATTGATTTATTTTACGGATTTTATGATTAATGCACAGATGCCTCATTTGCTCAATAAAT		2574
Qy	1190	TTTTTATAA		1198
Db	2575	TTATTTTAA		2583

```

RESULT 8
US-09-627-650B-7
; Sequence 7, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-7

```

Query Match 41.1%; Score 3097.5; DB 4; Length 2508;

Best Local Similarity 36.5%; Pred. No. 1.6e-136; Matches 779; Conservative 0; Mismatches 409; Indels 949; Gaps 120;		
QY	1	GATCTGGGTGC-----TGC--CA-----GGAAGCAAAATTCGGAA 37
Db	364	GATATGGACTTCACATTAGACTTCTACATGGCTCAACGTCGCAAGACCTCGACTAGCC 423
QY	38	GTTAATGGTTTGA--GTGATTTT-----AAATC-----CTTGTGGGGAGAG-- 81
Db	424	TTCGAAGTCTTGATTTGGGACTTTCCAAAGAAATCGACTCATTACCGTCGGAGTAGAC 483
QY	82	-----CCCGCC-----TCTCCCC----- 94
Db	484	TACCTGGATAGACTGTGGAAACCCGACACGTTCTTCCCAATGAAAGAAATCATTTCTC 543
QY	95	-----GGTATCAGCG--TTCCTCANT--CTTGAATC----- 123
Db	544	CATTGGACACACACTAACTGCTTCTTCTGATCGAGGGTGTGAAACGGTTTAACT 603
QY	124	-----CGGGCTCCGGCT--CTTCGGCT--CA-----GACCAG-----CCGAGGAA 163
Db	604	AGTCAAGATTAACAGTCACTGCAAGCTGTCATGACCTGAAAGCTGTTCCTCAATGGA- 662
QY	164	GCCTGTTTCAATTAAGC--GGCTGTGAACG--CCAGGGC--GGCGGGGC----- 212
Db	663	-----CTCTCAACACTGTAACTGAAATTTGAAGTACGGGTAGGACGAAAGATATCGAC 720
QY	213	-----GGGCGGAGGGCGCCATTTTGAATAAGAGGC-----GTGCTTTC 253
Db	721	TACTATTGGGGAAGAAGCGGACTGATTTGGAGATAACGGGTGTCAGTTTGAATCCTTC 780
QY	254	CAG-----GCAG-----GCTCTA--TAAGTGA-----CCG-----CCGG 281
Db	781	CAGTTGCGGAGTTTCAGCAACAGCTGTATTTTGTGATACAACTAAGACCGGACCTCA 840
QY	282	GCAG-----CGTGGC--GGTTGCAAGT-----C 305
Db	841	TCAGGAAATACGTACGCTGGCGTGGAGTAATATTGGTTGCAAAATATGGGCTTCTAC 900
QY	306	ACTGT--AGCG--GACTTC-----TTTTGGTTTCTTT----- 334
Db	901	ACTATGAACATCGTCAATCCATCCATCTGTGATCGTCAACCATATTCTGGGTATCATTTGG 960
QY	335	-----CTC-----TTTGGG-----GCAC----- 347
Db	961	TTGAATCGAAGCTTCGCGGCTCGAGTTGGATTGGGTGACTACTGTCTCACAATG 1020
QY	348	-----CTCTGACTC--ACTCCCA-----GCATGAAGCGCT-----GAGCCCG 385
Db	1021	ACBACTCTGATCACTACACCAATAATTCGATGCCCAAAAGTGTCTTATGTCAAGGTCG 1080
QY	386	G--TGCGGGC-----TGCTACGAGGCGGTGCTGCT-----GTGG--GAACGC--- 427
Db	1081	GATGTGTTCTTAATTTTGTTCGTATGTTGTTTCGATGTTTC--GCCTCGTGTCTCGAGTACCCAT 1139
QY	428	-----AGTCTGGCCATCGCCCGG-----GCCGAGGAAAG--GCC-- 461
Db	1140	AGTATCTCTACATGAATAACGACTGGTCTCCGACGGGAAACGAAAGAAAGCCGCGA 1199
QY	462	-----CGGCAGC-----TGAGAGCGCT-----GAGCTGTGAGACGAC-----A 497
Db	1200	ACAACAGCAGGAAACGAGATGCCAATGTTCAACGCGAGCGGAGGCGCCCAATAATA 1259
QY	498	TGAACCACTGCT--ACTCCC-----GCCTGCGGGA-----ACTGGTAC--- 533
Db	1260	TAATCCATTGATGGAGATCCGAAATAATTTGATTTGCCGAGATTCCAAATGATCAACA 1319
QY	534	CCG-----GAGTCCCGAGGCACTCA--GGTTAGCAGGTGGAATCCTACAGCGGT 585
Db	1320	TCACAGTCTGTACAGAGCGGCGCATACGTATGCGCGG-----CTCCATTGCGCGG 1374
QY	586	C-----ATCGACT-----ACATCT-----CGA-----CCTGCA-----GGTAGTCT 618

Db	1375	CGGAAAAGGCTTCCAAGACATGCTGCCAAGATGGACGCTGCAAAAATCGATAAGCTT 1434
QY	619	GGCGAGCCAG---CCC-----CTGG-----CTGG----- 636
Db	1435	AGCGATACGCTTTCCCAATGTCTTCTCTATCTTCAATATAGTACTGTTGTATATG 1494
QY	637	-----ACCC-----CCTGATGCCCCAC---CTTCCATC 664
Db	1495	AAATATCTAAGCTTAAACTCTGCGCAAGATCCAGAGAACACAAAGTGGAGAGATC 1554
QY	665	CA-----GACAGCCGAGCTCG-----CTCCGGAAC--- 689
Db	1555	CACATGATCGTATTTCGACGGCGGAAATCGAGTACAAATGCTACGTCGAAGGAGCGGAA 1614
QY	690	---TT-----GTCTATCCCAACGACAAAAG--GAGCTTTTTCG---CACT--G 726
Db	1615	TTGTTTCGACAGCGGTCAAGCGCGACGCGAATCGAATCTGCTGAGTTATAAATCTATA 1674
QY	727	ACTGCGCGTGTCTGACACCT--CCAGAAC-----GCAGG-- 760
Db	1675	AATCTGCAAAAGCGGACACTTGCACGACTTTCATCGGACCTACTCTCTGCTACGGGT 1734
QY	761	-----TGCTGGC-----CCGTTCTG 777
Db	1735	TAGTTTCATATTTGATCGGACAGCGGCTTCTACTTTCTTCAAAATATTTTCCCTGCCAG 1794
QY	778	CC-----TGGGACCC---CGGGAAC-----CT---CTCCTGCCGG 806
Db	1795	CCTCGTCTAGTTTTATCATGGATCTCATTTCTGATCAATCGTACTCGCGCGCTTCGCG 1854
QY	807	AAGCGG-----GACGCGAGG---ATGGGCCCAACTTC--GCCCTG--CCCACTTG 851
Db	1855	AACCTAATCGGTACGATGAGGTGCTCACTGAGACTCATCTTATGACCGAACCAATCG 1914
QY	852	AC---TTCACCA-----AATCC-----CT 867
Db	1915	ACGTTCTCCACCACTGCTCTATGTAAGCGGTTGATGTATTTCTCTCGTTTCTGTATCT 1974
QY	868	TC-----CTGG-----AGACTAAACCTGG-----TGCTCAGGAGCGAAGACTG 906
Db	1975	TCTGTTATATCTGGGTTGATCGAGTACGCTGTGTGCTACTCTCAAAAAGAGAACGA 2034
QY	907	TGAATTT---GTAGCTGAAG--AGCCAGAGCT---AGCTCTG--GCCACCAG--- 949
Db	2035	GGATCGTCGGAG 2094
QY	950	-----CTGGCGACGTCACCTT-----GCTCCACCC-----C 976
Db	2095	TCCTCAGCAGCTCGGCTTGCCGAATGCACATGCACGCGGCTCCAACCTCGATCATCG 2154
QY	977	CACCCCGAGTCTTAAGTCTTTTCAGAG-----CGTGGAGGTGGAAGGA 1023
Db	2155	CGTCATCAAGCAGTCGAATCGAATCTGTGTGAGTCAAGTCAATTGACAT--CGTCAGCC 2213
QY	1024	ETGGC-----TGCTC-----TCCAAACTATSCCAAGGCGG---CGGCAGA 1060
Db	2214	GTGCGCGTTCCTCTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2273
QY	1061	GCTGGTCTTCTGCTCT--CC-----TTGGAGAA---AGGTT--CTGTTGC-----CCTG 1102
Db	2274	ACAAATCCAGCGTCTGCGGTATATAGTGAACGAGGAGTACCGTTGCGATGCTCCAG 2333
QY	1103	A--TTATGAA--CTCTAT-----AATAGAGTATATAGTT--T 1135
Db	2334	ACCTTCAATTAATCTCAATCCAACTTCCTCATCTTTCATTTTCCATTTTCGAATATCTCTTTCT 2393
QY	1136	TGTAC-----CTTTTTTAC-----AGGAAGTGAATTTCTGTAAC-----AAT 1173
Db	2394	TGCACAGAGCGTTTTTCTGTTTTTTTTTTTATTTATTTTACCGATTTTTAGATAAT 2453
QY	1174	GC---GATG-----TATATTAACCTTTTATAA 1198
Db	2454	GCACAGATGCTCATGCTCAATAAATTTATTTTAA 2490

RESULT 9

US-09-436-063C-7
 ; Sequence 7, Application US/09436063C
 ; Patent No. 6407210
 ; GENERAL INFORMATION:
 ; APPLICANT: Jorgensen, Erik
 ; APPLICANT: Bamber, Bruce
 ; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
 ; TITLE OF INVENTION: Methods Related Thereto
 ; FILE REFERENCE: P-1095corrected
 ; CURRENT APPLICATION NUMBER: US/09/436.063C
 ; CURRENT FILING DATE: 1999-11-08
 ; PRIOR APPLICATION NUMBER: 60/107727
 ; PRIOR FILING DATE: 1998-11-09
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 2508
 ; TYPE: PRT
 ; ORGANISM: Caenorhabditis elegans
 US-09-436-063C-7

Query Match 41.1%; Score 3097.5; DB 4; Length 2508;
 Best Local Similarity 36.5%; Pred. No. 1.6e-136; Indels 949; Gaps 120;
 Matches 779; Conservative 0; Mismatches 409;

QY	1	CATCTGGGGTGC-----TGC--CA-----GGAAAAAGCAAAATTCGGAA	37
DB	364	GATATGGACTTCACATTAGACTTCTACATGGGTCAAACGTGGCAAGACCCTCGACTAGCC	423
QY	38	GTAAATGGTTTGA---GTGATTTT---AAATC-----CTTGCTGCGGAGAGG--	81
DB	424	TTGGAAGTCTTGATTTGGGACTTTCCAAAGAAATCGACTTACCTTACCGTCGGAGTAGAC	483
QY	82	-----CCGCGC-----TCFCCCC-----	94
DB	484	TACCTGGTAGACTGTGAAACCCGACACGTTCTTCCCAATGAAGAAGAAATCATTTTC	543
QY	95	GGTATCAGGC--TTCTCATTTCTTTGAATC-----	123
DB	544	CACITGGCAACACACATAACTGTTCTTCTGATCGAGGGTGTATGAACGGTTTATPACT	603
QY	124	---CGCGCTCCGGGT---CTTCGGCT---CA---GACCAG-----CGGAGGAA	163
DB	604	AGTCAAGATTAACAGTCACTGACAGCTGTCATGACGACCTGAAGCTGTTCCTCCAAATGA-	662
QY	164	GCCTGTTTGCAATTTAAGC--GGGCTGTGAACGC--CCAGGGCC--GGCGGGGGC-----	212
DB	663	---CTCTCAACACTGTAAACTGAAATTTGAAAGTACGGTACGGTACGAGACGAAAGATATCGAC	720
QY	213	-----GGGCGCGAGCGGGCCATTTTGAATGAAGGCG-----GTGCGCTTC	253
DB	721	TACTATTGGGGGAAGACGGGACTGATTTGGAGATAACGGCTGTCAAGTTTGATACCTTC	780
QY	254	CAG-----GCAG-----GCTCTA-TAAGTGA-----CCG---CGCG	281
DB	781	CAGTTGCCGCAATTCACCAACGCTGTATTTTGTGAATACAACTAAAGCGGAGACCTCA	840
QY	282	GGAG-----CGTGGC---CGTTGCAGT-----	305
DB	841	TCAGGAAATACGTACGCTCGCGCTGGGCTGGAAGTAAATATTTGTTTGAATATGGGCTTCTAC	900
QY	306	ACTGT-AGCG-GACTTC-----TTTTGTTTCTTT-----	334
DB	901	ACTATGAATCGTCAATCCCATCCATCCTGATCGTCACCATATCTTGGGTATCATTTTGG	960
QY	335	-----CTC-----TTTGGG-----GCAC-----	347
DB	961	TTGAATCGAGAAGCTTCGCGGCTCGAGTTGGATTTGGGTGTGACTACTGTGCTCACAATG	1020
QY	348	-----CTCTGGACTC-ACTCCCCA-----GCATGAAGCGCT-----GAGCCCG	385

DB	1021	ACAACCTCTGATCACTACAAACCAATAATTCGATGCCAAAGTGTCTTATGTCAAGGGTCTG	1080
QY	386	G-TGCGCGGC-----TGCTACGAGCGGTGTGCTGCCT-----GTCC-GAAGCG---	427
DB	1081	GATGTTTCTTAATTTTGTTCCTTAATGTTATTC-GCCTGTTGCTCGAGTAGCCAT	1139
QY	428	-----AGTCTGGCCATCGCCCGG---GCCGAGGGAAGG---GCC---A	461
DB	1140	AGTATCTTACATGAATAAACGACTGCTCTCGACGGGAAAAACGAAGAAGCCGCCGA	1199
QY	462	-----CGGCGC---TGAGGAGCGCT-----GAGCTTGTCTGGAGCAGC-----A	497
DB	1200	ACAACAGCAGCGAAACGAGATGCCAATGTTCAACCGGAGCCGGAAGGCGGCCCAATAA	1259
QY	498	TCGAACCACTGCT---ACTCCC-----GCCTGCGGGA-----ACTGGTAC-----	533
DB	1260	TAATCCATTGATGGAGATCCCAAGAAATTTGATTGCCGAGCATTCCTCAATGATCAACA	1319
QY	534	-CCG-----GAGTCCGAGAGGCATCA---GCTTAGCCAGGTGGAAATCTTACAGCGGT	585
DB	1320	TCCACGTCTTGTACAGACGGCGCACATACGCTATGGCCGG-----CTCCATTCGCGGG	1374
QY	586	C-----ATCGACT-----ACATTCT-----CGA-----CCTGCA-----GGTAGTCT	618
DB	1375	CCGAAAAGGCTTCCAAAGACATGCTGCCAACGATGGAGCGCTGCAAAATCGAATAGTT	1434
QY	619	GGCCGAGCCAG---CCC-----ACCC-----CTGTATGCTACTGCTTGTATATG	1494
DB	1435	AGCCGATACGGTTTCCCATTTGCTTCTCTATCTTCAATATAGTCTACTGCTTGTATATG	1494
QY	637	-----ACCC-----GACAGCCGAGCTCG-----CTCGGGAAC-----	689
DB	1495	AAATATCTAAGCTTAACTCTGCGGCAAGATCCAGGAGAGCAGCAAGTGGCAGGATC	1554
QY	665	CA-----GACAGCCGAGCTCG-----CTCGGGAAC-----CTCGGGAAC-----	689
DB	1555	CACGTATGGTATTTCGACGGCGGAAATCGAGTACAAATGGTGTACGTGCAAGGAGCCGAA	1614
QY	690	---TT-----GTCATCTCAACGACAAAG-GAGCTTTTC-----CACT-G	726
DB	1615	TTGTTGACAGCGGTCAAGCGCGACGCGAAATCGAACTGTGAGTTATAATTCCTAA	1674
QY	727	ACTCGCCGCTGTCTGACACCT-CCAGAAC-----GCAGG-760	760
DB	1675	AACTGCCAAAAACGACACTTGGCAGCACTTCATCGGGGACCTACTCTCGTCTACGGGT	1734
QY	761	-----TGCTGGCG-----CCGTTCTG777	777
DB	1735	TAGTTTCATATTGATCGGACAGCGGTTTCTACTTTCTTCAAAATATTTTCCCTGCGCAG	1794
QY	778	CC-----TGGGACCC---CGGGAAC-----CT---CTCCTGCGCG806	806
DB	1795	CCTCGTGTAGTTTATCATGATCTCATCTGATCAATCGTGAATCGGCGCTTCGCG	1854
QY	807	AAGCG-----GACGCGAGG---ATGGGCCCCAACTTC-GCCCTG-CCCACTTG	851
DB	1855	AACCCCTAATCGGTACGATGACGCTGCTCACTGAGACTCATCTTATGACCGGAACCAATCG	1914
QY	852	AC-----TTCACCA-----AATCC-----CT867	867
DB	1915	AGTCTTCCACCAGTTGCTATGTAAGCGGTTGATGTATCTCTCGGTTCTGTATCT	1974
QY	868	TC-----CTGG-----AGACTAAACCTGG-----TGCTCAGAGCGAAGAGCTG	906
DB	1975	TCTGTTTATCTGCGGTTGATCGAGTACGCTGTGTGCTACTCAAAAAGAGAAGCA	2034
QY	907	TGAACCT---GTAGCCTGAAG-AGCCAGAGCT---AGCTCTG---GCCACCAAG---	949
DB	2035	GATCTGTGGAAG	2094
QY	950	-CTGGGCGAGCTCACCT-----GCTCCACG-----C976	976

Db 2095 TCCTACGAGCTCCGGCTTGGCAATGACATGCAACCGGCTCCAACTCGATCATCGC 2154
 Qy 977 CACCCCAAGTCTTAAGGCTTTTTCAGAG-----CGTGAGGTGTGGAAGGA 1023
 Db 2155 CGTCATCAAGCAGTGAATCGATTTCTGTCAGTCACAGTCACATTCACAT-CGTACGCC 2213
 Qy 1024 GTGGC-----TGCTC-----TCCAACTATGCCAGCGG---CGGCAGA 1060
 Db 2214 GTCCCGCTTCTCTGTGTTTCATCTGTTTCAACACTCTCTTCTGCTGATTCATCTGT 2273
 Qy 1061 GTGTGCTTCTGTCT-CC-----TTGGAGAA---AGTTT-CTGTTC-----CCTG 1102
 Db 2274 ACAAAATCCAGGCTCTGCCGTATATAGTGAACAGCGGTGACCGTTGCGATGCTCCAG 2333
 Qy 1103 A-TTATGAA-CTCTAT-----CAACTTCCCTCATCTTTTCCATTTGATACTCTTTTCT 2393
 Db 2334 ACCTTCATTAATCTCAATCAACTTCCCTCATCTTTTCCATTTGATACTCTTTTCT 2393
 Qy 1136 TGTAC-----CTTTTTC-----AGGAGGTGACTTTCTGTAAAC-----AAT 1173
 Db 2394 TGCACAGAGCCTTTTTCGTTTTTTTATGATTTATTTTACGGATTTTGTAGATAAT 2453
 Qy 1174 GC---GATG-----TATATTAAACTTTTATAA 1198
 Db 2454 GCACAGTGCCTCATTTGCTCAATAAATTTATTTAA 2490

RESULT 10
 US-09-627-650B-3
 ; Sequence 3, Application US/09627650B
 ; Patent No. 6406872
 ; GENERAL INFORMATION:
 ; APPLICANT: Baber, Bruce
 ; APPLICANT: Jorgensen, Erik
 ; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
 ; TITLE OF INVENTION: Methods Related Thereto
 ; FILE REFERENCE: 21101.000903
 ; CURRENT APPLICATION NUMBER: US/09/627,650B
 ; PRIORITY FILING DATE: 2000-07-28
 ; PRIOR APPLICATION NUMBER: 09/436,063
 ; PRIOR FILING DATE: 1999-11-08
 ; PRIOR APPLICATION NUMBER: 60/107,727
 ; PRIOR FILING DATE: 1998-11-09
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 2544
 ; TYPE: PRT
 ; ORGANISM: Caenorhabditis elegans
 US-09-627-650B-3

Query Match 41.1%; Score 3097.5; DB 4; Length 2544;
 Best local Similarity 36.1%; Pred. No. 1.6e-136;
 Matches 784; Conservative 0; Mismatches 404; Indels 985; Gaps 125;
 Qy 1 GATCTGGGTGC-----TGC-CA-----GGAAAAAGCAAAATCTGGAA 37
 Db 364 GATATGGAATTCATATGACTTCTACATGCTCAACGTCGCAAGACCCCTCGACTAGCC 423
 Qy 38 GTTAATGGTTTGA---GTGATTTT---AATC-----CTGTGGCGGAGAG-- 81
 Db 424 TTCGGAAGTCTTGATTTGGGACTTTCCAAAGAAATCGACTCACTTACCGTCGGAGTAGC 483
 Qy 82 -----CCCGCC-----TCTCCCC----- 94
 Db 484 TACCTGGATAGACTGTGGAACCCGACACGTTCTTCCCAATGAAAGAAATCATTTCTC 543
 Qy 95 -----GGTATCAGCGC-TTCCTCAT-TCTTTGAATC----- 123
 Db 544 CACTTGGCAACACACATAACTCGTTCCTTCGTATCGAGGGGTGATGGAACGGTTTATCT 603
 Qy 124 ---CGCGGTCCGCGGT---CTTCGGCGT---CA---GACCAAG-----CCGAGGAA 163

Db 604 AGTCAAGATTAAACAGTCACTGCAACGCTGTCCAAATGGAAGCTGTTTCCCAATGGA- 662
 Qy 164 GCCTGTTTGAATTAAGC-GGCTGTGAACGC-CCAGGCGC--GGCGGGGC----- 212
 Db 663 --CTCTCAACACTGTAAACTGGAATTAAGAGCTACGGGTACGAGACGAAAGATATCGAC 720
 Qy 213 ---GGGCGGAGCGCGGCCCATTTTGAATAAAGAGGC-----GTGCCTTC 253
 Db 721 TACTATTGGGGAGAGAGCGGACTGATTTGGAGATAACGCTGTCAAGTTTGTACTCTTC 780
 Qy 254 CAG-----GAG-----GCTCTA-TAAGTGA-----CCG---CCGCG 281
 Db 781 CAGTTGCGCGAGTTTTCAGCCAAACGCTGTATTTTGTGAATACAACTAAAGCGGAGACTCA 840
 Qy 282 GCGAG-----CGTGGCG--GCCTTCAGGT-----C 305
 Db 841 TCAGAAATAATACGTACGCTGCGCTGCGCTGGAAGTAAATATTTGTTGAAATATGGCTTCTAC 900
 Qy 306 ACTGT-AGCG-GACTTC-----TTCGTTTTCCTTT----- 334
 Db 901 ACTATGAACATCGTCACTCCCATCCATCCTGATCGTCAACATATCTTGGGTATCATTTTGG 960
 Qy 335 -----CTC-----TTTGGG-----GCAC----- 347
 Db 961 TTGAATCGAGAAAGCTTCGCGGCTCGAGTTGGATTGGGTGAGTACTGTGCTCACAATG 1020
 Qy 348 ---CTCTGACTC-ACTCCCA-----GCATGAAGCGCT-----GAGCCCG 385
 Db 1021 ACAACTCTGATCACTACACCAATTAATTCGATGCCAAAGTGTCTTATGTCAAGGGTCTG 1080
 Qy 386 G-TGCGCGC-----TGCTACGAGCGGTGTGCTGCT-----GTGG-GAAGCG-- 427
 Db 1081 GATGTGTTTCTTAATTTTGTGTAATGTTTC-GCCTCGTGTCTCGAGTACGCCAT 1139
 Qy 428 -----AGTCTGGCCATCGCCCGG-----GCCGAGGAGG--GCC-- 461
 Db 1140 AGTATCTTACATGAATAAACGACTGGTCTCCGACGGGAAAAACGAAGAAAGCCCGGA 1199
 Qy 462 ---CGGCGC---TGAGGAGCGCT-----GAGCTTCTGCTGAC----- 493
 Db 1200 ACAACAGCAGCGAAACGAGATGCCAATGTTCAACGCGAGCGCCGAGCCCAATAATAA 1259
 Qy 494 ---GACATGAC-----CACTGCTACTC-----CCG-----CCTGC 521
 Db 1260 TGCTGACTGTGACTTTTCCGCGACACAATCTCTTATGAATCCATGATGAGAGATCCAGA 1319
 Qy 522 GGGAACTGTTACCGG---GAGTCCCGAGAGGCA-----CTCAGCTTAGC-CAGGTGGAAA 572
 Db 1320 AAATTTGATGTCGCGGACGATTCACATGATGCAACATCCACGCTTGTACAGACGCGC 1379
 Qy 573 TCCTACAGCGGTCTCATGACTACATTTCTGACTGTGAGGTAG-----TCTGGCCGCGC- 626
 Db 1380 ACATACGCTATGGC--CGGTCCATTCGCG--CGGCGGAAAAAGGTTCCAGACATGCT 1435
 Qy 627 -CAGCCCTGACCCCT-CTGA-----TGGCC-----CCCA-----CCT 657
 Db 1436 GCCAACGATGACGCTGCAAAAATCGAATAGTTCAGCCGATACGCTTTCCCATTTGCTTT 1495
 Qy 658 TC---CCATCCAGACAGCC-----GAGCTCGCT---CCGGAATCTGTC-- 694
 Db 1496 TCTCTATCTTCATATAGTCTACTGGTGTATGAATATCTAAGTTAAATCTGTCGG 1555
 Qy 695 ---ATCTCC-----AAGCAAAAAG-AGCTTTTGCACCTGA-----C-TGGCC-- 734
 Db 1556 ACAAGATCCAGGAGAACGACAAAGTGGCAGCAGAT-CCACTGATGCGTATTCGACGCGCGA 1614
 Qy 735 -----GTGTCC-----TGACACCC---TCCAGAAAGC 757
 Db 1615 AATCGAGTACAAATGTTGTAGCTGGAAGGAGCCGAATTTGTCACAGCGGTCAAGCGGA 1674
 Qy 758 AG-----GTGCTGGCG-----CCC---GTTCTGCTG---GACCC-- 787
 Db 1675 CGCGAACATCGAACTGTGCGAGTTATAAATTCATAAATCTGCAAAAGCGGACACTGC 1734

```

788 788 -----CGGGAAC---CTCTC---CTGC-----CGAAG 809
Db 1735 CAGCACTTCACTGGGACCTACTCTCTGTCTACGGGTAGTTTCATATTTGATCGGACAG 1794
QY 810 -----CGGACGG-----CAGGGAT 824
Db 1795 CGGCTTCTACTTCTTCAATAATTTTCCCTGCAGCCTCTGTCTAGTTTATTCATGGAT 1854
QY 825 -----GGGCC---CAACTTCG---CCC-----T 842
Db 1855 CTATTCGTGATCAATCGTACTCGGCGCTTTCGGAACCTTAATCGGTACGATGCGT 1914
QY 843 GCCCACT-----TGAC-----TTACCA-----860
Db 1915 GCTCACTGAGACTCATCTTATGACCGGAACCAATCGACGCTTTCACCAAGTTGCCTATGT 1974
QY 861 -----AATCC-----CTTC-----CTGC-----AGA 876
Db 1975 AAAAGCGTGTATCTCTCGGTCTCTCTATCTTCTGGTTATATCTGCGTTGATCGA 2034
QY 877 CTAACCTGG-----TGCTCAGGAGCGAAGGACTGTGAATTT-----GTAGCCTGAAG-A 925
Db 2035 GTACGCTCTGTGCTCTACTCAAAAAGAAAGACGAGATCGTCGGAGAAGAGAGAA 2094
QY 926 GCCAGAGCT-----AGCTCTG---GCCACGAG-----CTGGGCGAGTCACTCT-----966
Db 2095 GACGAGCATAAACCTCTCTCGCGCACACCTGATATTTTCACGACGTCGCGCTTCCCGA 2154
QY 967 -----GCTCCACC-----CCACCCCAAGTCTTAAGGTCTTTT 1000
Db 2155 ATGCACATGCAACGCGCTCAACCTCGATCATCGCGTCAATCAAGCAGTCGATCGAT 2214
QY 1001 CAGAG-----CGTGGAGGTGTGAAGGAGTGGC-----TGCTC-----1033
Db 2215 CTGTGTGAGTCACAGTCACATGACAT---CGTCAGCGCTGCGGCTTTCTCTTTGTTTCA 2273
QY 1034 -----TCCAACTATCCAAAGCGG---CGGAGAGCTGTCTCTGCTCT---CC-----1078
Db 2274 TCTGTTCACACACTCTCTCTGCTGATTTCTGCTGATTTCTGCTGATTAATCAAGCTCTGCGGTATA 2333
QY 1079 TTGAGAA-----AGGTT---CTGTTC-----CTGA---TTTATGAA---CTCTAT-----1117
Db 2334 TTAGTGAACAGAGGAGGAGCGGTGCGATGCTCCAGACCTTCATTAATCTCAATCCAAC 2393
QY 1118 -----AATAGATATATAGTT---TTGTAC-----CTTTTATAC-----1149
Db 2394 TCCTCATCATTTCCATTTCAATTTACGATTTTATAGATTAATGCACAGATGCTCATTTGCTCAAT 2453
QY 1150 AGGAAGGTGACTTTCTGTAAC-----AATGC---GATG-----TATATT 1185
Db 2454 TTTTATTTGATTTATTTTACGATTTTATAGATTAATGCACAGATGCTCATTTGCTCAAT 2513
QY 1186 AAACCTTTTATAA 1198
Db 2514 AAATTTATTTAA 2526

```

```

RESULT 11
US-09-436-063C-3
; Sequence 3, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-3

Query Match 41.1%; Score 3097.5; DB 4; Length 2544;
Best Local Similarity 36.1%; Pred. No. 1.6e-136;
Matches 784; Conservative 0; Mismatches 404; Indels 985; Gaps 125;

QY 1 GATCTGGGGTGC-----TGC---CA-----GGAAAAAGCAAATTTCTGGAA 37
Db 364 GATATGGACTTCACTTAGACTTCTACATGGTCAAGCGTGGCAGACCTCGACTAGCC 423.
QY 38 GTTAATGTTTGA---GTGATTTT---AATC-----CTTGTCGGCGGAGAGG-- 81
Db 424 TTCGAAGTCTTGATTTGGGACTTTCCAAAGAAATCGACTCACTTACCGTCGAGTAGAC 483
QY 82 -----CCGCGC-----TCTCCCG-----94
Db 484 TACCTGGATAGACTGTGTGAAACCCGACACGTTCTTCCAAATGAAAGAAATCTTCTTC 543
QY 95 -----GGTATCAGCGC---TTCCTCAT---CTTTGAATC-----123
Db 544 CACTTGGCAACACACATAACTCGTTCTCTCGTATCGAGGGTGTGAAACGGTTTATACT 603
QY 124 ---CCGCGCTCGCGGT---CTTGGCGT---CA---GACGAG-----CCGAGGAA 163
Db 604 AGTCAAGATTAACAGTCACTGCAACGCTTCCAAATGGACCTGTGAAAGCTGTTCCCAATGGA- 662
QY 164 GCTCTTTGCAATTTAAGC---GGGCTGTGAAGC---CCAGGGGC---GGCGGGGGC-----212
Db 663 ---CTCTCAACTGTAAACTGAAATTAAGAGTACGGGTACGAGACGAAAGATATCGAC 720
QY 213 -----GGGGCCAGCGCGGCCATTTTGAATAAGAGGC-----GTGCTTTC 253
Db 721 TACTATTGGGGGAAGAGCGGACTGATTGGAGATAACGGCTGTCAAGTTTGATACCTTTC 780
QY 254 CAG-----GCAG-----GCTCTA---TAAGTGA-----CCG---CCGCG 281
Db 781 CAGTTGCGCAGTTTCAGCCACGCTGTATTTTGTGATACACTAAGCCGAGACCTCA 840
QY 282 CGGAG-----CGTCCGC---CGCTTGCAGGT-----C 305
Db 841 TCAGAAATAACGTCAGCGCTGCGCTGGAAGTAATATTGGTTTCAAAATATGGGCTTCTAC 900
QY 306 ACTGT---AGCG---GACTTC-----TTTTGGTTTCTTT-----334
Db 901 ACTATGAACATCTCATCCATCCATCCTGATCGTCACCATATCTTGGGTATCATTTTGG 960
QY 335 -----CTC-----TTTGGG-----GCAC-----347
Db 961 TTGAATCGAAGACTTGCCTGCTCGAGTTGGGTGTGACTACTGTGCTCAAAATG 1020
QY 348 ---CTCTGGACTC---ACTCCCA-----GCATGAGGCGCT-----GAGCCCG 385
Db 1021 ACAACTCTGATCACTACAAACCAATTAATTCGATGCCAAAGTGTCTTATGTCAGGGTCTG 1080
QY 386 G-TGGCGGGC-----TGCTACGAGCGGTGTGCTGCT-----GTG-GAAGC-----427
Db 1081 GATGTGTTTCTTAAATTTTGTGTTTGTGTAATGTTATTC-GCTCGTTGCTCGAGTACGCCAT 1139
QY 428 -----AGTCTGGCCATCCCGCGG---GCCGAGGAGG--GCC--461
Db 1140 AGTATCCTTACATGAATAAACGACTGGTCTCGAGCGGGAACCAAGAAAGACCGCCGA 1199
QY 462 ---CGGAGC---TGAGGAGCGCT-----GAGCTTGTCTGGAC-----493
Db 1200 ACAACAGCAGCGAAGAGATGCCAATGTTCAACGCGAGCGCCGAGCCCAATAATAA 1259
QY 494 ---GACATGAAC-----CACTGTACTTC-----CCG-----CCTGC 521

```

Db 1260 TGCTGACTTGTTGCGGACACAAATTCCTCTATGAATCCATGTATGGAGATCCCGAGA 1319
 Qy 522 GGGAACTGTATCCCG---GAGTCCCGAGAGCA-----CTCAGCTTAGC-CAGGTGMAA 572
 Db 1320 AAATTGTGATTGCGGACGATTCGAATGATGCAACATCCAGCTCTGTTCACAGACGGCGC 1379
 Qy 573 TCTACAGCGCTCATCTGACTACATTCCTGACTCGAGTAG-----TCCTGGCGGAGC- 626
 Db 1380 ACATAGCTATGCG---CGGCTCATTCGCG---CGGCGGAAAAGGCTTCCAGACATGCT 1435
 Qy 627 -CAGCCCTGAGCCCG---CTGA-----TGCGC-----CCGA-----CCT 657
 Db 1436 GCCAACGATGCGCGCTGCAAAAATCGATAAGCTTAGCGGATACGGTTTCCCATGTGCTT 1495
 Qy 658 TC-----CCATCCAGACAGC-----GAGCTCGCT---CGGAACCTGTGCT-- 694
 Db 1496 TCTCTATCTTCAATATAGTCTACTGTTGTATATGAATATCTAAGCTTAACTGCTCGG 1555
 Qy 695 ----ATCTCC-----AACGACAAAGG-AGCTTTTCCACTGA-----C-TGGGCC-- 734
 Db 1556 ACAAGATCCAGGAGACGACAAAGTGGCAGCAGAT-CACTGATGGTATTGACGGCGGA 1614
 Qy 735 -----GTGTC-----TGACAC-----TCCAGAACGC 757
 Db 1615 AATCGAGTCAATGCTGTACGTCGAGAGCGGAATGTTTCGACAGCGGTCAAGCGCGA 1674
 Qy 758 AG-----GTGCTGGCG-----CCC---GTTCTGCCTG---GGACCC----- 787
 Db 1675 CGGACATCGAATGCTGAGTTATATAATCTCAATAAATCTGCCAAAACGACACTTGC 1734
 Qy 788 -----CGGAAAC---CTCTC---CTGC-----CGGAAG 809
 Db 1735 CAGCACTTCATCGGACCTACTCTGCTCAGGGTTAGTTTCATATTGATCGCGAGAG 1794
 Qy 810 -----CGGACCGG-----CCGACCGG-----CAGGGAT 824
 Db 1795 CGGCTTCTACTTCTTCAATATTTTCCCTGCCAGCCTCCGTCGAGTATGATGAT 1854
 Qy 825 -----GGGCC---CAACTTCG-----CCC-----T 842
 Db 1855 CTCATTCGTGATCAATCGTACTCGCGGCTTCGCGAACCTTAATCGGTACGATGACGT 1914
 Qy 843 GCCCACT-----TGAC-----TTACCA----- 860
 Db 1915 GCTCCTGAGACTCATCTTATGACCGGAACCAATCGACGCTTCCACCACTTGCTATGT 1974
 Qy 861 -----AATCC-----CTTC-----CTGG-----AGA 876
 Db 1975 AAAAGCGGTGATGATTAATCTCGGTTTCTGCTATCTTCTGTTATATCTGCGGTGATCGA 2034
 Qy 877 CTAAACCTGG-----TGCTCAGAGCGAAGGACTGTGAATTT---GTAGCTGAAG-A 925
 Db 2035 GTACGCTGTGTTGCTTACTCAAAAGAAAGACGAGGATCGTCGGAAGAGAGAGAA 2094
 Qy 926 GCCAGAGCT-----AGCTCTG---GCCACAG-----CTGGGCGAGCTCACCT- 966
 Db 2095 GACGGAGCATAAACCTGCTCGCGGACACCTGATATTTCTCAGACGCTCGCGCTTGCCGA 2154
 Qy 967 -----GCTCCACC-----CCACCCCAAGTTCTAAGTCTTTT 1000
 Db 2155 ATGCACATGCAACCGGCTCAACCTCGATCATCGCGCTCATCAACGATCGAATGAT 2214
 Qy 1001 CAGAG-----CGTGGAGGTGTGGAAGGAGTGGC-----TGCTC----- 1033
 Db 2215 CTGTGTCAGTCACAGTCACATGACAT-CGTACGCGTGGCGGTTTCTCTGTGTTTCA 2273
 Qy 1034 -----TCCAACTATGCCAGCGG---CGGAGAGCTGCTTCTGCTCT-CC----- 1078
 Db 2274 TCTTGTTCACACTCTCTTCTGGCTGATTCTACTGTACAAATCAAGCGTCTGCCGTATA 2333
 Qy 1079 TTGAGAA-----AGGTT---CTGTTC-----CCTGA---TTTATGAA-CTCTAT----- 1117
 Db 2334 TTAGTGAACAGAGGGTGACCGTTGGCATGCTCCAGACCTTTCATTAATCTCAATCCAAT 2393

Qy 1118 -----AATAGAGTATAGGTT--TTGTAC-----CTTTTTC- 1149
 Db 2394 TCTCATCATTTTCCATTTTCGAATATCTCTTTTTCGACAGAGCCCTTTTTCGTTT 2453
 Qy 1150 AGGAAGGTGACTTTCGTAAAC-----AATGC---GATG-----TATATT 1185
 Db 2454 TTTTATTGATTATTTTACGGATTTTAGATAATGACAGATGCTCATTCCTCAAT 2513
 Qy 1186 AACTTTTATAA 1198
 Db 2514 AATTTATTTTAA 2526

RESULT 12

US-08-977-767-3
 ; Sequence 3, Application US/089777767
 ; Patent No. 5972684
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Yue, Henry
 ; APPLICANT: Greenwald, Sara
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/977,767
 ; FILING DATE: Herewith
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0423 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-855-0555
 ; TELEFAX: 650-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1345 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 1532042
 US-08-977-767-3

Query Match 40.7%; Score 3073.5; DB 2; Length 1345;
 Best Local Similarity 46.3%; Pred. No. 1.3e-135;
 Matches 663; Conservative 0; Mismatches 423; Indels 345; Gaps 86;
 Qy 4 CTGGGGTCTCCAGGAAAAGCAATTTCTGGAAGTTAATGGTTTTCAGTGAATTTTAAA 63
 Db 8 CTGAGGGTT---CGGAAGGAAAG---GGAAGCAAGCGAGGAGG----- 52
 Qy 64 TCCTTGCTGGCGAGAGCGCCCTCTCCCGGTATCAGCGCTTCCTCAT-TCTTTGAAT 122

Db 53 ----AGGAGAGAGAGAGGGGTGGGGTCCAG--ATCCCGCTGCTGTAACCTGGGAA 107
Qy 123 CCGCGCTCC-----GGGTCCTTCGCG-GTACAGACAGCCGCGAGGAGCTGTT 170
Db 108 CTGGACTCCCTCCAGGAGAGCG--AAGGCTGGAGTCTTG--GGAGGATTAGGT 162
Qy 171 TGCATTTTAAAGCG-----GCTGTGAACCCAGAGCGCGCGGGGCG-----GGGCG 217
Db 163 TCCACAAGAGGGGAGAGCTAGAG-----AGGGCCGGCGGGAGCATTTGGGAAGGGT 217
Qy 218 CGAGGGCGGGCAATTTGMAATAAGAGCG-GTGCT--TCCAGCGAGCTC-----TATA 268
Db 218 CACCCGAGGCTCTTGGGATGGGGCTGCACCTCGTCTGAG--GCCCTCGGGTGT 275
Qy 269 AGTGAC--GCCGCG--GGAGCGTGGGGGCTTGA--GGTCACTGTAGGAGT---- 318
Db 276 GGTACTCTGGCGCGCACTGGAGCG--GCACGTCACTCGGACCC--GCACCTGACCGA 331
Qy 319 --TCCTTTGT-TTTC-----TTCTCTTTGGGCACT-----CTGACTCACTCCC--- 363
Db 332 GGACTGTGAGCTACAGGATATCTCAGGGAATCTGTGCCAGGAC---CTCCCTTC 388
Qy 364 -----CAGCATGAAGCGCTGAGCC-CGGTGGCGGC--TGTACGAG 404
Db 389 TGGGGCTGGTGAACGAGCGCTGAGTCTCTGTGCGGTGGAGCGCTCAGAGCCCGTGG 448
Qy 405 CGGTGTC-TG-----CCTCGGACGAGTCTG-GCCATC-----GCCGGGGCGG 450
Db 449 ATGTGACGTGAAGAGGGTCTCTATGACCCCTTCTGTGCCCTCTGAGACTCAGACCG 508
Qy 451 AGGGAAGGGCCCGCAGCTGAGGAGCGCTGAGCTT-GCTGAGCAGCATGAACCACTGT 509
Db 509 GGGAGAG---AAGCCTCCGTGAAC-TCTGTACAACACCGTGGCCATG---TCTCCT 559
Qy 510 ACTCCCGCTGGGGAAGTGTACCGGAGTCCGAGAGGAGCTAGCTAGCCAGGTGG 569
Db 560 --TCCTGCTGCG-----CCCCGG--CCTGTGG-----TTAATGTGTGGGGG 600
Qy 570 A-AATCTACAGCGCTGATCACTCATTTCTCGAC--CTGC-----AGGTAGTCTTGGCC 622
Db 601 ACCTCTTTATAG--CACGACTCAGTGAATGCGGCTGCTATTTGGGAGCAGGAG 658
Qy 623 GAGCAGCCCTGAGCCCTGTATGGCCCC-----CAC-----CTTCCCAT 663
Db 659 GAGCTGCTCT--GAACACAGATCAACCATCAGGGTTCTCTGCTGAGGTGCACTCAT 716
Qy 664 CCA-GACAGCGAGCTCGCTCCGGAATTTGTCTATCTCCACGACAAAGAGCTTTTGGC 722
Db 717 CCATTTCAACAGAACTCTATGGGA-----ACCTCAGTGGCGCCACAGGGCC----CC 768
Qy 723 ACTGACTCGCGCTGTCTGACACC-----TCCAGACGAGGT-GCTGGCGCCGTTTC--- 775
Db 769 AATGGCTGGCCATTCTCAGCCTCTTTGTCAATGTGGCTGGTAGCTCAAAACCGTTCTC 828
Qy 776 ---TGCCCT-----GGGACCCCGGAAC--TCTCCT--GCCGGAAGCGG--A 814
Db 829 AGAGCCTCTTAACCTGACATCATCCCGCATCTCTATAGATGATGCTACTT 888
Qy 815 CGGAGGG-ATGGGCCC-----CACTTCGCGCTTC-CGACTGACTTCAACCAATCCCTT 868
Db 889 CTTCAAGACCTGAGCCTGGAGCTCCTGTGCCCGAGTCTTTGGCTTCATCA-----CCTA 944
Qy 869 CTTGG-----AG-ACTAAGCTGGTCTCAGGAGCGAAGGACTGTGAATGTGAGCC 919
Db 945 TCAGGGTCTCTAGACCCCAACC--TGCTUG-----AGACTGT-TACCTGATTC 994
Qy 920 TGA-----AGAGCCAGAGCTAGTCTTGGCCACAGCTGGGCGACGTACCGTCTCCAC 974
Db 995 TCAATTGACGGGCCCTCAATATCACCTCCCTCCAGCTG-----CAT-TCCCTCGCACTTC 1048
Qy 975 C---CCA-----CCCCAGTTCTAAGTCTTTTTCAGAGCTGGAGGTGT-GGAGG-AGT 1025
Db 1049 TGAGCCAGAAATCTCTCCGGTCCCA--GTC-TTCAGAGCTCAGCCGTAACGGCCGCCCC 1105

Qy 1026 GGCTGCTCTCCAAACTATGCCAAGGC-----GGCGGAGAGCTGCTTCTTGCTCTCTT 1081
Db 1106 TGCAGCCCTTCCGCG-----ACAGGGCTTTGAGGGC--AAACAGACCCCGGCCACCC-- 1157
Qy 1082 GAGAAAGGTTT-----TGTG-----CCCTGATTAT 1108
Db 1158 GAGAGGGGTGGGAGGAGGCCCACTACCGCTGCTGATGTGGATGATGTCCTCCCATGCTCT 1217
Qy 1109 GA---ACTCTATAATAGATATATAGTTTGTACCTTTTAC-----AGGAA 1154
Db 1218 GAGACTCCCATCAGGAGTCCGCTGCTT-TCCCAAGCTTCCCAACCGAGGGGAGGG 1276
Qy 1155 GGTCACTTCTGTAACTAATCGAT-----GTATATTAACTTTTATAAAA 1200
Db 1277 AGTGACCCCAAAAGCTATTAAAGGACAGATATCTTCTCTGTAATA 1327

RESULT 13

US-09-627-650B-5
; Sequence 5, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,083
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1917
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-5

Query Match 39.4%; Score 2971.5; DB 4; Length 1917;

Best Local Similarity 39.7%; Pred. No. 8.6e-131; Indels 671; Gaps 108;
Matches 729; Conservative 0; Mismatches 435;

Qy 3 TCTGGGCT---GCTGCCAGGAAAAAGCAAAATTC---GGAAGTT-----AATGG 45
Db 100 TCTGTCTCCGCTAAACG--ATTGGCTCATTTCTTCTCGGCTCTTGTCACTAACGAATCG 157
Qy 46 TTTT-----GAGTGATTTT---TAAATCTTGC-----TGG-----CGGA--- 77
Db 158 CTTTTCACACAGAGTGATATAGGAGAAAAACCTCCCAACATTGGCTCACACCCGATTA 217
Qy 78 -----GAGG--CCGCGCTCTCC----- 92
Db 218 TGAATCTTCTGCTGCTCTGCTGCTCTCTCTGCTAGTGTAGACGAGAAAGAAAGAAG 277
Qy 93 ---CCGGTATC-----AGCGCTTC--TCATCTTTGAATCCCGGCTCGCGGTCTTC 141
Db 278 CTCCATTTCTCGAAGATGGCTCGTCCATTCACATTCATCTATCTGCTCTCTCCG--ACATC 335
Qy 142 GCGCTC-----AGAC-CAGCGGAGGAAGC----- 165
Db 336 TGTGTCTACATGTGGTGTGTGACACAGGATGAGGACTCACATATCAACACTCACTCTCT 395
Qy 166 ---CTGTTTGAAT---TAAAGCGGT-----GTGAACGCCAGGCGCGGCGG 208
Db 396 CATCAGTTCTCGATAGACTCAGGAATCGCACTACTTATGATAAAGATTACGGCC--CAG 453
Qy 209 GGGCGGGGCCGAGGCGGCCCAT-TTTGAATAAAGAG--GCCTGCTTCCA---GGCAGG 261
Db 454 GTATGTTGAAGGCCAGTCTGAGTGGATACGATACACGTTTCTTCAATCTCTGCACT 513

336 CATCAGTTCTCGATAGACTCAGCAATCGCACTACTTATGATAAAAGATTACGGCC--CAG 453

209 GGGCGGGCCGAGGGGGCCAT-TTTGAATAAAGAG--GCGTGCCTTCCA-----GGCAGG 261

454 GTATGGTGAAGACCCAGTCGAGGTGGAAATTACGATACACGTTTCTTCAATCTCTGCAAT 513

262 CTCCTATAGT-----GAC--CGCC-----GCGCGAGCGTGGCGGTTTGCAGGTC 305

514 TTC-AGAAGTTGATATGGACTTCACATTTAGACTTCTACATGGCT-CAACAGTGGCAAGAC 571

306 ACTGTAGCGGAGCTCTCTTTTGGTTTTCTCTCTTTGGGCG--ACCTCTG-----GACTCA 358

572 CCT---CGACTAGCCTTCGGAAGTCTT--GATTTGGGACTTTTCAAGAAATTCGACTCA 625

359 CTCCTCCAGC-ATGAAGC-----GCTGAG--CCCGGTGCG-----CGG-- 393

626 CTTACCGTCGGAGTAGACTACCTGGAATAGACTGTGGAACCCGACACAGTTCTTCCCAAT 685

394 -----CTGCTACGAGCG-----GTGTGCTCCCTGTC----- 420

686 GAAAGAAATCATTCTTCCACTTGGCAACACACATAACTCGT-TCCTTCTGATCGAGG 744

421 -----GGAACG-----CAGTC-----TGGCCATCGC-----CCGGGCGCAGGGAAGG 458

745 TGATGGAACGGTTTATACAGTCAAAAGATTAACAGTCACTGCAACGTGTCC-----AATG 799

459 GCCCGCAGCTG-----AGGAGCGCGTGAGCT--TGCTGGACGACATGAACCACTGC 508

800 GACCTGAAGCTGTTCCCAATGACTCTCAACACTGTAAACTGGAAATTGAAGCTACGCG 859

509 TAC---TCC-CGCC-----TG-----CGGAA-----CTGTACCCGGAGTC- 541

860 TACAGTATCTCGACATTATGTACGTGTGCGACGAGAAGAAGTCCGTGTCCACCGAGTCT 919

542 -----CCG-----AGAG-----CACTC-----AGCTT 559

920 TATGAGTTCGCGAGTTTGTACTTCAAGTCTATCAAGTCTGTCGTAATCATAGCNAAGCTT 979

560 AG-CCAGGTGGAAA-----TGCTA--CAGCGCGTCACTCGACTAC 595

980 AGTTCAGGAAATATCCCGCCTTGTGTTCTTCTCTATTCAAGCGTAACATCGGCTTC 1039

596 ATTCTGCAC-----CTGCAAGTAGTCTCTGGCG-AGCCAGCCCTTGACCCCTCGA 645

1040 TACATCATCAAAATATATCTACCATCTGCTCCTGATTGTGCTGATCTCATGGGTATCTTTT 1099

646 TG-----GCCCC-----CACCTTCG-----CATCCAG--ACAGCCGAGCTCGCT 682

1100 TGGTTGAGCCGCGATGCGACCCGCGAAGAGTTGCTCTCGAGTCAACCACCTGTGCTCA 1159

683 CCG-----GAACCTGTGATCTC--CAACGACAAAG----- 711

1160 ATGACTACTTTGATGACCATGACTAATAGTTCATGCCAAAGTGTCATATGTGAAAAGT 1219

712 ---GAGCTTTTGCA-----CTGACTC--GGCGGTGCTCTG-ACACCTCCAGA--ACGCA 758

1220 ATCGATATATTTCTAGTGTCTGCTTCATGATGGTATCTGTTCACCTTCTAGATAGCC 1279

759 G--GT-----GCTGGCGCCGCTTC-TGCGTGG----- 782

1280 GCCGTGCGATACATCAACAAACGATGAAGCTTTCGAGCCGAAAGAAATCTCGAATG 1339

783 --GACCCCGGGAAC-----CTCTCTGCGGAAGCCGAGCGGAGGATGGCC 829

1340 CTGACCCCTTTTACCACATCTTGAATCTTCTCTCC--AAACGTACTCTA--TCCGTTC 1395

830 C-----CAAACCTTCGCC-----CTGCCCACTTGACT--TCACCAATC-----CCTT 868

1396 CTCGTATTTCAACAACACCACTGACCGCCCGCTTTTACTGTGTCCACCGATCAACGCTCAA 1455

869 CTGGAGACTAAACCTGGTGTCTAGGAGCGAAGACTGTGACT-----TGTAGCTCTAAG 924

1456 CTTGTACATT---CCGGAGTCGC---AGCGACACAGGATTTTCTCAAATG-AGGATGAC 1508

QY	925	AGCC---AGAGCTAGCTCTGGCCACCAGCT-GGCGACGTCACCCCTGCTCCACCCACC	980
Db	1509	TGCGGAATGAACTTACT---CCA-ATGCTCGACGGAGTAACTC-ACAAGCATCCGTAT	1562
QY	981	CCCAAGTTCTAAAGTCTTTTCAGAGCGTGGAGTGT---GGAA-----GGAG-TGGCT---	1029
Db	1563	TTCTGTATCAGACGGCTGTAATATCCGATCAGAGTTCGAAGATTCTGCGTGGCTCC	1622
QY	1030	-----GCTC-----TCCAAAATA-----TGC-----CA	1047
Db	1623	GACCATCCAACATTGACAAGTACTCAGCTCACTGTGCCATCTATTTTGTGCTCTTCA	1682
QY	1048	A-GCGGGCGCAGAGCTGG---TCTTCTGGT-----CTCCTTGGAG---AAAGTTTCT	1093
Db	1683	ACGTGCGCTACTGGGCTACTTTCATCGGAGAGCCAGATTCAGGAAGACCAAGGAACA	1742
QY	1094	GT---TGCCTCATTTATCAAC-----TCTATAATAGATAT---	1127
Db	1743	GTCAATTCTTAATTTCTGATCACACCACTCCTCATCTCATTTCTATTGTAGCCTTTTT	1802
QY	1128	-----ATAGGTTTG---TACCTTTTACAGGAAGTGACTTTCGTAAACNA-----	1172
Db	1803	TTTCGAATCATTTCTGAATATCTCTTATCTTCTCAAGATGGCATCGCTAAACCACCCC	1862
QY	1173	TGCGATCTATATAAAGT---TTTTATAAAGTT	1203
Db	1863	CACCTAGCACAAAAATGCTCTTCTAGAAATTT	1897
RESULT 15			
US-09-627-650B-1			
; Sequence 1, Application US/09627650B			
; Patent No. 6406872			
; GENERAL INFORMATION:			
; APPLICANT: Bamber, Bruce			
; APPLICANT: Jorgensen, Erik			
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and			
; TITLE OF INVENTION: Methods Related Thereto			
; FILE REFERENCE: 21101.0009U3			
; CURRENT APPLICATION NUMBER: US/09/627,650B			
; CURRENT FILING DATE: 2000-07-28			
; PRIOR APPLICATION NUMBER: 09/436,063			
; PRIOR FILING DATE: 1999-11-08			
; PRIOR APPLICATION NUMBER: 60/107,727			
; PRIOR FILING DATE: 1998-11-09			
; NUMBER OF SEQ ID NOS: 50			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 1			
; LENGTH: 1652			
; TYPE: PRT			
; ORGANISM: Caenorhabditis elegans			
US-09-627-650B-1			
Query Match 39.2%; Score 2954.5; DB 4; Length 1652;			
Best Local Similarity 41.6%; Pred. No. 4.7e-130;			
Matches 680; Conservative 0; Mismatches 458; Indels 497; Gaps 98			
QY	1	GATCTGGGTCTG--CCAGGAAAAACAAATTCG--GAAGTTAATGGTTTGAGTGAT	56
Db	40	GATCT--TCTGCTGCTCTG---CTGCTCTTTCTGCTGTAGTTCA-GACGAGAAGAAG	92
QY	57	TTTAAATCCTTGTGGCGGAGAGCCCG--CCTCTCC-----CCTCTCC-----	92
Db	93	AAGAAGCTCCATTCGAGAAATGGCTCGTCCATTACACTTATGCTACTCTCTCCGCA	152
QY	93	-----CCGG-----TATCAGCGCT-----TC	108
Db	153	CATCTGTCTCATGTGGTTGTGCACAGATGAGGACTCACATATCAACTCAACTC	212
QY	109	CTC-----ATTCT---TTGAATC-----CGCGCTCC-----CGGTTC	138
Db	213	CTCTCATGATTTCTCGATAGATCTCAAGATCGC-ACTACTTATGATAAAGATTACGGCC	271

QY 139 -----TTGGCGGTGACAGCAGCGGAG-----GAAGCGTGT-----TGC 173
 Db 272 CAGGTATGTGAAAGCCAGTGCAGCTTGGAAATACGATACACGTTTCTTCAAICTCTGC 331
 QY 174 AATTAAAGCGG--GCTGTGAAC--GCCA--GGGC-----CGCG--GGGCGGG 216
 Db 332 AGTTTCAGAAAGTGTATGAGCTTCAATAGACTTTACATGCTTCAAAAGTGGCAAGA 391
 QY 217 -CGAGCGGGGCC-----AATTTGAATAAGAGCGGTGCCTTCCAGCAGGCTCTATA 268
 Db 392 CCTCGACTAGCTTCGGAAGTCTTGATT--TGGACTTTCAAAAGAAATCGACTCACTT 449
 QY 269 AGTGACCGCGCGCGAGCGTGCAGCGT-----TGCAGGTCACTGTAGCGGACTT--CTT 322
 Db 450 -----ACCGTGGAGTAGAC--TACCTGGATAGACTGTGGAACCCGACAGCTTCTCCAA 504
 QY 323 TTGGT-----TTTCTTTCTTTTGGGCGACCTCTGGACTCACTTC-----CCAGC 367
 Db 505 ATGAAGAAGAAATCACTTCTTCCACTTGGCAACACACATAACTCGTTCTTCTGATCGAGG 564
 QY 368 ATGAAGCGGTG-----AGCC--CGGTGCGCG--CTGCTAGGAGC-----GG--- 407
 Db 565 GTGATGAACGGTTTACTAGTAGTCAAGATTAACGTCACCTGCAAGCTGTCCAATGGACC 624
 QY 408 TG-TGCTG---CC--TG---TCGGAACGCACT---CTGGC----- 435
 Db 625 TGAAGCTGTTCCTCAATGAGCTCTCAACACTGTAACTGGAATTAAGCTATGCGTATT 684
 QY 436 CATCGCCCGGGCGGAG-----GGAAGGCC-----CGGAGCTG 470
 Db 685 CGAGCGCGGAAATCGAGTACAAATGTGTACGTGCGAAGGAGCGGAATGTTGACAGAGCG 744
 QY 471 AGGA--GCGGCTGAG-----CTTGTGAGCAGACATGAAC---CACTGCTA--CTCCCGCCTG 520
 Db 745 TCAAGCGCGAGCGGAACATGAACT--GTGAGTTATAATTAATTAATTAATTAATTAATTA 803
 QY 521 CGGGAACCTGTACCGGAGTCCGAGAGGCACTGAGCTTAGC--CAGGTGGAATCCTTAC 578
 Db 804 CGGACACT--TGCCAGCACTTCACTCGGGGACCTACTCTGCTACGCGGTAGTTTCATAT 861
 QY 579 ---AGCGGTCATCGATCACTTCTGACCTGAGGTAGTCTGCGGCGAGCCAGCC--- 631
 Db 862 TTGATCGGAGCGGGCTTC--TACT--TTTCTCAAATATTTT---CCCTGCGAGCCTCGT 916
 QY 632 -----CTGGACCCCTGATGCGCC-----CACCTTC--CCATCC-- 665
 Db 917 CGTAGTTTATCATGGATCTCATCTGATCAATCGTGACTCGGCGCTTCGGCAACCT 976
 QY 666 -----AGACAGCCGAGCTCGT-----CGGACTTGTCTCATCTCCA 701
 Db 977 AATCGGTACGATGACGGTGTCTCACTGAGACTCATCTTATGACCGGAACCAATCGAGTCT 1036
 QY 702 ACGACAAAAGAGCTTTT-----GCCACTGA-----CTCGGCGGTGCTCTGACACCT 748
 Db 1037 TC--CACCAGTTGCTATGTAAGCCGTTGATGATTTCTCGGT---TTCGTCTATCT 1090
 QY 749 CCAG--AAGCGAGGTGCTG-----GGGCGGTTCTGCTGGGACCC----- 787
 Db 1091 TCTGTTTACTGGCGTTGATCGAGTAGCGCTGTGTTGCCT---ACTCAAAAAGAGAA 1147
 QY 788 CGGGAACCTCTCTCGCGAAGCGGAGCGGAGGATGG--GCCCAACTTCGCGCTGCC-- 845
 Db 1148 CGAGATC-----GTGCG--AGAAGAGAGAGAGAGCGGACATAAACCT--GCTCGCGCG 1199
 QY 846 -CACTTGA-----CTTCAACAATC--CCTTCTGAGACTAACT--GGTGTCTAG--- 893
 Db 1200 ACACCTGATATTCTTTCAGAGCTCGCGCTTCGGAATGACATGCAACGCGGCTCCAAC 1259
 QY 894 --GAGCGAGGACTGTGAGCTGTGAA--GAGCCAGAGCTAG--CTCTGGC----- 943
 Db 1260 TCGATCATCG---CGGTCTCAAGCAATC--GAATCGATTCTGTGCTGCTGACAGTCACTT 1316

QY 944 -----CACAGCTGGCGGAG-----TCACCGTGTCTCCACCCACCC----- 982
 Db 1317 GACATGCTGAGCGTGCAGGTTTCTCTTGTTCATCTTGTTCACACTCTCTTCTGG 1376
 QY 983 CAAAGTTCTAAGGT--CTTTTCAGAGCGT--GGAGGTGTGGAAG-----GAGTGGCTGC 1031
 Db 1377 CTGATTCTACTGTACAAATCCAAGCGTCTGCCGTATATTAGTGAACACGAGGGTGACCGT 1436
 QY 1032 T-----CTCCAAAC-----TATGC--CAAGGCGGCGGAGAGCTGTCTTCTGGTCTCC 1078
 Db 1437 TGGATGCTCCAGACCTTCAATCTCAATCCAATTC-----CTCATCAT-----TTTCC 1488
 QY 1079 TTGAGAGAAAGGTTCTG---TTGCCCTGATTTATGAACCTCTATAATAGAGTATATAGTT 1134
 Db 1489 ATTTGGAATATCTTTTCTTCTGCACAGA-----AGC-CTTTTTCGTTTTTTTT--TA 1539
 QY 1135 TTGTACCTTTTACAG-----GAAGGTGACTTTCTGTAACAATGCGATGATATATA 1187
 Db 1540 TTGATTTATTTTACGGATTTTATAGATAATGACAGATGCTCATTC-----TCAATATA 1595
 QY 1188 ACTTTTATAAAGT 1202
 Db 1596 ATTTATTTTAATGCT 1610

Search completed: May 7, 2004, 15:11:36
 Job time : 58 secs

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:10:05 ; Search time 54 Seconds
(without alignments)

(without alignments)
6183.570 Million cell updates/sec

Title: X691111

perfect score: 7544
Sequence: 1 GATCTGGGGTGCTGCCAGGA.....TTAAACTTTTATAAAAGTT 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

```

1: /cgn2_6/prodata/2/pubpaa/us07_PUBCOMB.pcp.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/prodata/2/pubpaa/us06_NEW_PUB.pcp.*
4: /cgn2_6/prodata/2/pubpaa/us06_PUBCOMB.pcp.*
5: /cgn2_6/prodata/2/pubpaa/us07_NEW_PUB.pcp.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/prodata/2/pubpaa/us09A_PUBCOMB.pcp.*
9: /cgn2_6/prodata/2/pubpaa/us09B_PUBCOMB.pcp.*
10: /cgn2_6/prodata/2/pubpaa/us09C_PUBCOMB.pcp.*
11: /cgn2_6/prodata/2/pubpaa/us09D_PUBCOMB.pcp.*
12: /cgn2_6/prodata/2/pubpaa/us09_NEW_PUB.pcp.*
13: /cgn2_6/prodata/2/pubpaa/us10A_PUBCOMB.pcp.*
14: /cgn2_6/prodata/2/pubpaa/us10C_PUBCOMB.pcp.*
15: /cgn2_6/prodata/2/pubpaa/us10_NEW_PUB.pcp.*
16: /cgn2_6/prodata/2/pubpaa/us16_NEW_PUB.pcp.*
17: /cgn2_6/prodata/2/pubpaa/us60_PUBCOMB.pcp.*
18: /cgn2_6/prodata/2/pubpaa/us60_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	3593	47.6	2236	12	US-10-142-426-381		Sequence 381, App
2	3593	47.6	2236	14	US-10-123-155-381		Sequence 381, App
3	3593	47.6	2236	14	US-10-146-731-381		Sequence 381, App
4	3593	47.6	2236	14	US-10-140-472-381		Sequence 381, App
5	3593	47.6	2236	14	US-10-141-761-381		Sequence 381, App
6	3593	47.6	2236	14	US-10-142-885-381		Sequence 381, App
7	3593	47.6	2236	14	US-10-158-790-381		Sequence 381, App
8	3593	47.6	2236	15	US-10-137-871-381		Sequence 381, App
9	3593	47.6	2236	15	US-10-140-523-381		Sequence 381, App
10	3593	47.6	2236	15	US-10-141-756-381		Sequence 381, App
11	3593	47.6	2236	15	US-10-141-759-381		Sequence 381, App
12	3593	47.6	2236	15	US-10-140-805-381		Sequence 381, App
13	3593	47.6	2236	15	US-10-140-864-381		Sequence 381, App
14	3576.5	47.4	3721	12	US-10-142-426-543		Sequence 543, App
15	3576.5	47.4	3721	14	US-10-123-155-543		Sequence 543, App

Query Match 47.6%; Score 3593; DB 12; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

ALIGNMENTS

RESULT 1

```

US-10-142-426-381
; Sequence 381; US/10-142426
; Publication NO. US2004004833A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Rustin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE P
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C224
; CURRENT APPLICATION NUMBER: US/10/142,426
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See file Wrapper or
; NUMBER OF SEQ. ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-426-381

```

[illegible]

QY 1 GATCTGGGCTGCTCCAGGAAAGCAAAATTC--TGAAGTTAATGTTTGA-----GTG 54
Db 191 GCGCGCGGGAACCGG-----GCGAATCTCTCGCGCTCAAAACACCTGATCCATA 244
QY 55 ATTTTAAATCTTCTGCTGCG-----GAGAGGCGCGCTCTCCCGGTATCAGC 103
Db 245 AAACATTCATCTC-CGCGCGCGCGCTGCGAGCGCGCGCGAGTCCGCGCGCGCC 303
QY 104 GCTTCTCATTTTGAATCC-GCG-GCTCGCGGTTCTTGGGTC---AGACCAGCGG 158
Db 304 GC-CCTCGCGCTGTGCGCTGCGCGCTGCG--CACCGCGCGCGAGCGCCAGCCAG 359
QY 159 AGGAAGCCTGTTTGAATTTAAGCGGCTGTG-AAAG-CGCAGGCGC-GGCGGCGCGG 215
Db 360 AGCGGCGG--AGCG-----GAGCGCGCGAGCGCTGCTCCGCGCGCGCGCGCG 413
QY 216 GCGGAGGCGG-----GCCATTTTGAATAAGA--GGC-----GTGCTTCCA 255
Db 414 GCGGAGCGGCGGCGGCG-----TGATGCGGACCGCGCGCGCGGAGACGCGCGCG 469
QY 256 GCGAGGCTCTATAAGTGACCGCGCG-----GCGAGCGTGGG 292
Db 470 CGAAACGACTTTCAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
QY 293 C-----GC-GTTGAGGTCA-----CTGTAGC-----GGACT-TCTTTTGGTT--T 329
Db 530 CTGAGGAGCGCGCTGTGTG-CATGGGTGTGTGTGCGGCGCTGCGAGGCTTGGCAGCC 588
QY 330 TCTTTCTCT-----TTGGG---GCACCTCTGGAATCA-----CT-----CCCCG 366
Db 589 CCATCCCGAGTGTCTGCTGATGCTACAAATGAGCCCAAGGTGACGAAAGTGCGCCCG 648
QY 367 CA-----TGAAGGCGCTGAGCGCG-GTGGCGGCTGCTACGAGCGGTGTG-CTGCTG- 418
Db 649 CAGGCGCTGCAG--GCTGTGCGCGGTGGGCAATCCCTGTGCGAGCCAGCGCAATTTCTG 706
QY 419 TCGGAA-----CGCA---GTCTG--GCCATCGCGCGCG---GCCGAG----- 452
Db 707 ACGGCAACCGCATCTGCGATGTGCGAGCTGCGAGCTTCCGTGCGCGCGCAACTCACCA 766
QY 453 -----GGAAG-----GGCGG-----GCAGCTG----- 470
Db 767 TCCTGTGCTGACTGGAATGTGTGCGCGAATGTGATGCGGCTGCGCTTCACTTGGCGTGG 826
QY 471 -----AGGAGCGGCTG---AGC---TTGC-----TTGG----- 491
Db 827 CCTCTGAGGAGCTGAGCTGAGCTCAGCGAATATGACAGCTCGGTCTGTGAGCCCTGCCA 886
QY 492 -----AGCATGAAACAC-----TGCTACTCC-----C-GCCTGCGGAGAC 527
Db 887 CATTCCAGCGCTGGCGCGCTACACAGCTGACCTGAGACCGCGCTGCGGCTGCGAGAGC 946
QY 528 TGGTACCGGAG-----TCC-----GCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1005
Db 947 TGG-GCCCGGCGCTGTTCCGCGGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1055
QY 544 GAG-----AGGCACT-----CAGCTT-----AGCCAG----- 566
Db 1006 GCGCTGAGGAGCTGCTGATGACACTTCCGAGCTTGGGACCTGCGGACCTGCGGACCTTTC 1065
QY 567 -----TGAAA-----TCCTAC-----AGCGCTCATC-----GACT--ACATTC 601
Db 1066 CTGACCGGCAACCGCATCTCCAGCGTGCAGGCGCGCTTCCGCTGCGGCTGCGAGCGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
Db 1126 GACGCTCTCTACTGACCAAGACCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1185
QY 618 --TGCGG-----AGCCAGC----- 630
Db 1186 CTGCGCGCTCATGACACTCTATCTGTGTTGCGCAACAATCTATCAGCGCTGCGCCTGAG 1245

QY 631 -CCTTGACCCCTGATGCGCC--CC--ACCT-----TC-----CATCC-----A 666
Db 1246 GCGCTGCGCGCTGCGTGCCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1305
QY 667 GACAGCGGAGCTCG-----CTCCGGAATCTGT-----CATCTCCAAACGACAA 709
Db 1306 GACTGCGCGGACGCGCACTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1365
QY 710 AGGAGCTTTTGGC-----ACTGACTCGCGCTGCTCT--GACACT-----CCA- 751
Db 1366 GTGCGCTGCGAGCTTCCGCAACGCTGCGCT--GGCGCTGACTCAAGCGCTGCTGCGCA 1424
QY 752 -GAAC-GCAGG--TGC-----TGCG-----GCGCGT-----TCTG--C--CTGGGA 784
Db 1425 TGACTGAGGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1484
QY 785 CCGCG-----GGAACCTCT-----CCTGCGCGAGCGG--GAC--GG-- 817
Db 1485 CACGATGAGGAGCGCTGCGGCTTCCCAAGTGTCTCCAGCGCAGATGCGCTGCAAGCG 1544
QY 818 --CAGGATGG-GCC-----CCAACTTCG----- 838
Db 1545 CTGAGTACTGAGCTTGAAGACCACTTCCGCGAGCAATGCGTGAAGGAGCGGTGCG 1604
QY 839 --CC-----CTGCGCACTTGAATTC-----ACCAATCTCTTCC-----T 871
Db 1605 GCGCGTGAAGCGCGCGCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1664
QY 872 GAGACTAAACCTGG-----TGCTCA-----GGAGCGAAGG-ACCTG 907
Db 1665 TGGGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1724
QY 908 G-----AATCTG-----TAGCTGAGAGCGCAG-----AGCTAGCTC 939
Db 1725 GCACACGAGGTTCCTGCTGCGGCGCTGCGCGA--GGCCAGGCTGTTCACGCAAGAC 1782
QY 940 TG-----GCCA-----CCAGCTGGG-----GAGC----- 959
Db 1783 CGCACCGCGACCACTGCGCTGCGCGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1842
QY 960 -----TCA-----CCTGCTC-----CCACCGCGCGCG----- 982
Db 1843 TCAGAGGCTCAGTGGCGCTTACCGCGCTCAGCTGAGCTCAGCTGAGCTCAGCTGAGCTG 1902
QY 983 CAAGTCTAAGGCTTTTC----- 1009
Db 1903 CTGTGCTGTGAGCAGTGTGTTGGCGCTTGGCGCTGCTGAGCGCGCGCGCGCGCGCG 1962
QY 1010 -----GAGGTGTG-----GAAGGAGTGTGCTCTCCAACTATGCCAAGCG-----GG 1053
Db 1963 GCAGCAGGTGTGTACATACGGGT--CT-CTCTCAGCG--CGCCAGCGCGCGCG 2017
QY 1054 CGGCGA-----GCTGCT--TCTGCT--CTCTGAGAAAGGTTCTTGTGCGCTG 1102
Db 2018 CG 2075
QY 1103 ATTTATGAACTCTA--TAATAGAGTATAGGTTTG-----TACCTTTTACAGGAG 1155
Db 2076 CACCCCGCTCCACCGCTATGTTTACAGGTTGCGCGCGCGCGCGCGCGCGCGCGCG 2135
QY 1156 GTGACTTCTGTAACAAATGAGTATATA---TTAACTTTTATAAAGTT 1203
Db 2136 CCGCTCCCAACCGAGATCGCGGTATATAGATATGCAATTTTATTACTT 2186

RESULT 2

US-10-123-155-381

; Sequence 381, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura

```
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3330RIC30
/ CURRENT APPLICATION NUMBER: US/10/123,155
/ CURRENT FILING DATE: 2002-04-15
/ Prior Application removed - See Palm or File Wrapper
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 381
/ LENGTH: 2236
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-10-123-155-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171; Indels 863; Gaps 125;
Matches 808; Conservative 0; Mismatches 360;

QY 1 GATCTGGGTGCTGCGCAGGAAAGCAATTC--TGAAAGTTAATGTTTGA-----GTG 54
DB 191 GCGCGCGGGGAAACCGG-----GCGGATTCCTCGCGGTCAAAACCACTGATCCCAT 244
QY 55 ATTTTAAATCTTGTCTGGG-----GAGAGCGCGCTCTCCCGGTATCAGC 103
DB 245 AAACATTCATCTC-CGCGCGCGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 303
QY 104 GCTTCTCTCATCTTTGAATCC-GCG-GCTCCGCGGTCTTGGCGTC---AGACCAAGCGG 158
DB 304 GC--CTTCGCGCTGTGGCGCTCGCGCGCTGCG--CACCGCGCGCGCGAGCCAGCCAG 359
QY 159 AGGAAGCTCTTTGAATTTAAGCGGTGTG-AAAG-CCAGGCGC-GCGCGGCGCGG 215
DB 360 AGCGGCGCG--AGCG-----GAGCGCGCGAGCTCTGTCGCGCGCGCGCGCGCGCG 413
QY 216 GCGGAGCGCG---GCCATTTGAATAAGA---GCG-----GTGCTTTCCA 255
DB 414 GCGTAGCGCGCGCGC---TGATGCGGACCGCGCGCGCGGAGACGCGCGCGCGCGC 469
QY 256 GCGAGGCTCTAATGTAACCGCGCG---GCGAGCGCGCGCGCGCGCGCGCGCGCGCG 292
DB 470 CGAAACGACTTTTCACTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
QY 293 C-----GC-GTTTCAGTCA-----CTGTAGC---GGACT-TCTTTTGGTT--T 329
DB 530 CTGAGGAGCGCGCTGCTGTG-CATGGGTGCTGTGGTGTGCGAGCGCTGCGAGGTGGCAG 588
QY 330 TCTTTCTCT-----TTGGG---GCACCTCTGACTCA-----CT-----CCCCAG 366
DB 589 CCATGCCCGAGTGTGCTGCTATGCTACAATGAGCCCAAGGTGACGACAAAGCTCCCGCC 648
QY 367 CA-----TGAGGCGCTGAGCGCG-GTGGCGGCTGTACGAGCGGTGTG-CTGCTGTG 418
DB 649 CAGGCGCTGCGAG--GCTGTGCGCGGTGGGATCCCTGTGTGCGAGCGCGCATCTTCTGCG 706
QY 419 TCGGAA-----CGCA-----GTCTG--GCCATGCGCGCGG---GCCGAG----- 452
DB 707 ACGGCAACCGCATCTGCGATGCGCAGCTGCGAGCTTCCGTGCGCTGCGCGCAACCTCACCA 766
QY 453 -----GGAAG-----GGCCCG-----GGAGCTG----- 470
```

```
DB 767 TCCTGTGGCTGCACTCGAATGTGTGCGCCGCAATTGATGGGTGCTGCTTCACTGGCGCTG 826
QY 471 -----AGAGCGCGTG-----AGC-----TTGC-----TGG----- 491
DB 827 CCTCTGTGAGCAGCTGAGCCTCAGCGATATGCAAGCTCCGGTGTGTGGACCTTGCCA 886
QY 492 -----ACGACATGAACCCAC-----TGCTACTTC-----C-GCTTCGGGGAAC 527
DB 887 CATTCACGCGCTGGCGCGCTACACACGCTGCACCTGGACCGCTCGGGCTCAGGAGC 946
QY 528 TGTATCCCGAG-----TCC----- 543
DB 947 TGG-GGCCGGGGCTGTTCGCGGGCTGGGTGCGCGCTGAGTACCTTACCTGAGGACAC 1005
QY 544 GAG-----AGCACT-----CAGCTT-----AGCCAGG----- 566
DB 1006 GCGCTGAGCACTGCTGATGACACTTCCGCGAGCTGGGCAACCTGACACACCTTCTTC 1065
QY 567 -----TGGAAA-----TCTTAC-----ACGGGTGATC-----GACT--ACATCTC 601
DB 1066 CTGCACGCAACCGCATCTCCAGGTGCGCGAGCGGCGCTTCGCTGGGTGACAGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
DB 1126 GACGCTCTCTACTGACACGACCGGTGCGCCATGTGACCGCGATGCTTCCGTGAC 1185
QY 618 -----TGCGCG-----AGCCAGC----- 630
DB 1186 CTTGGCGCTCATGACACTCTATCTTTTGCCAACTATATAGCGCTGCCACCTGAG 1245
QY 631 -CCCTGACACCGCTGATGGGCC--CC-ACT--TC-----CCATCC-----A 666
DB 1246 GCGCTGCGCGCTGCGTGCCTGCTGAGTACTGAGGCTCAACGACACCCCTGGGTGT 1305
QY 667 GACAGCGAGCTCG-----CTCGGAATTGT-----CATCTCAACGACAA 709
DB 1306 GACTGCGCGGACGCCCACTCTGCGGCTGGGTGAGAAATTCGCGGCTCTCTCTCGAG 1365
QY 710 AGGAGCTTTTGC-----ACTGACTCGGCGGTGCT--GACACCT-----CCA- 751
DB 1366 GTGCGCTGAGCTCCCGCAACGCTGCT--GGCGGTGACTCAAGCGCTAGTGCCAA 1424
QY 752 -GAAC-GCAGG--TGC-----TGGC-----GCCGT-----TCTG--C-CTGGGA 784
DB 1425 TGACTGCGAGGCTGCGTGTGGCGACCGGCGCTTACCATCCATCTGGACCGGAGCGG 1484
QY 785 CCGCG-----GGAACCTCT-----CCTGCGGAAAGCG--GAC--GG- 817
DB 1485 CACGATGAGGAGCGCTGGGCTTCCAAAGTGTGCGACCAATGCGGTGACAGGC 1544
QY 818 --CAGGATGG-GCC-----CCAATTCG----- 838
DB 1545 CTCAGTACTGAGCTTGAAGACAGCTTCGCGAGGCAATGCGTGAAGGAGCGGTGCC 1604
QY 839 --CC-----CTGCCCACTTGACTTC-----ACCAATCCCTTC-----T 871
DB 1605 GCCCGGTGACAGCGCGCGGCAACGCGCTTGGCGCCAGGCAATCAATGACTCACCTT 1664
QY 872 GGAGACTAAACCTTG-----TGCTCA-----GGAGCGAAGG-ACTGT 907
DB 1665 TGGGACTCTGCTGCTGCTGAGCGCCCGCTCACTGAGTGGCGCGCGGAGGTCCGA 1724
QY 908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGCTAGCTC 939
DB 1725 GCGACCAAGGTTCCCACTCGGCGCTCGCGCGA--GGCCAGGCTGTTCAGCAAGAAC 1782
QY 940 TG-----GSCA--CCAGCTGGG-----GAGC----- 959
DB 1783 CGCACCGGACCACTGCGTGTGGCGCAGGAGCGAGCGGGGTGGCGGAGTGGTGAC 1842
QY 960 -----TCA-----CCCTGCTC-----CCACCCACCC----- 982
DB 1843 TCAGAAGGCTCAGTGGCTTACCCAGCTCACCAGCTCAGCTGAGCCCTCACCCCTTGGGCTGGCG 1902
```

QY 983 CAAAGTTCTAAGGTCCTTTTC-----AGAGCGTG----- 1009
Db 1903 CTGTGCTGTGACAGTGTGTGGCCCTGCTGACCCCGAGCGGACACAGAGCGTGTCTCA 1962
QY 1010 -----GAGGTGTG-----GAGGAGTGGCTGTCTCCAACTATGCCAAGGC-----GG 1053
Db 1963 GCAGCCAGGTGTGTGTACATACGGGT--CT-CTCTCCACGC--CGCCAAAGCCAGCCGGG 2017
QY 1054 CGGCAG-----GCTGTCT--TCTGTCTCTCTCTGAGAGAGTTCTGTGCCCTG 1102
Db 2018 CGGCGAGCCGTGGGAGCGGAGGCGAGGTCTCTCTGATGGAG--CTGCGCGCGCG 2075
QY 1103 ATTATGAAGTCTA--TAATAGATATATAGTTTGG-----TACCTTTTATAGAGAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTTCGGCGGAGCGGTTTGTTCAGAGAG 2135
QY 1156 GTGACTTCTGTAACTATGATGATATA--TTAAACTTTTATATAAGTT 1203
Db 2136 CCGCTCCCAACCCAGATCGCGGTATATAGATATGATATTTTATTACTT 2186

RESULT 3

US-10-146-731-381
; Sequence 381, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Deanoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary B.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C23
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-146-731-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;
QY 1 CATCTGGGTCTGCCAGAAAGCAATTC--TGGAAGTAAATGTTTGA---GTG 54
Db 191 GCGCGGCGGGAAACCGG-----GCGAATTCCTCGCGGTCAACACCTGATCCATA 244
QY 55 ATTTTAAATCCTTGCTGGCG-----GAGAGCGCGGCTCTCCCGGTATCAGC 103
Db 245 AAACATTCATCTC-CGCGGCGCGCGCTCGAGAGCGCGCGCGAGTCCGCGCGCGCG 303
QY 104 GTTCTCTATCTTTTGAATCC-GCG-GCTCGCGGTCTTGGGCTC---AGACAGCGCG 158
Db 304 GC--CCTCGCGCTGTGCGCCTGCGCCCTGCG--CACCGCGCGCGCGAGCCAGCCAG 359

QY 159 AGAAGCCTGTTTGCATTTTAAGCGGGCTGTG--AAGC-CCCAGGGCC-GCGGGGGGGGG 215
Db 360 AGCGGGCGG--AGCG-----GAGCGCGCGAGCCTGTCTCCGCGCGCGCGCGCGCGG 413
QY 216 GCGAGAGCGG---GCCATTTTGAATAAAGA---GGC-----GTGCTTCCA 255
Db 414 GCGTAGCGGGGGCGCC---TGATCCGACCCCGCGCGGGGAGACGGCGCGCGCC 469
QY 256 GGCAGGCTCTATAGTGAACCGCGCG-----GCGAGCGTGTGCG 292
Db 470 CGAAACAGCTTTCAGTCCCGAGCGCGCGCGCCCAACCCCTACGATGAAGAGAGCGCTCCG 529
QY 293 C-----GC--GTTCAGGTCA-----CTGTAGC-----GGACT--TCITTTGGTT--T 329
Db 530 CTGAGGAGCGCGCTCTCTG--CATGGTCTGTGTGCTGAGGCTTGGCAGGTGGCAGCC 588
QY 330 TCTTTCTCT-----TTGGG---GCACCTCTGGAATCA-----CT--CCCCAG 366
Db 589 CCATGCCAGGTGCTGCTATGCTACATAGAGCCCAAGGTGACGACAAGCTGCGCCCGAG 648
QY 367 CA-----TGAGGGCTGAGCCG-GTGGCGGCTGTACGAGCGGTGTG-TGTGCTG- 418
Db 649 CAGGCGCTGAG--GCTGTGCGCGGTGGGCTATCCCTGTCTGAGCCAGCGATCTTCTGTC 706
QY 419 TCGGAA-----CGCA---GTCGTG--GCCATGCGCCCGG---GCCAG- 452
Db 707 ACGCAACCGCATCTCGCATGTGCGAGTCCAGCTTCCGTGCTGCGCAACCTCACCA 766
QY 453 -----GGAAG-----GGCCG-----GCAGCTG----- 470
Db 767 TCCTGTGCTGCACTCGAATGTGTGCGCGCAATGATGCGGCTGCTTCACTGGCCTGG 826
QY 471 -----AGGAGCGCTG-----AGC-----TTTG-----TGG----- 491
Db 827 CCTCTCTGAGCAGCTGGAACCTCAGCATATATGACAGCTCCGCTGTGTGACCTGCCA 886
QY 492 -----ACGACATGAACCAAC-----TGCTACTCC-----C-GCTCGGGAAC 527
Db 887 CATTCACGCGCTGGCGCGCTACACAGCTGACCTGAGCCGCTGCGGCGCTGAGGAGC 946
QY 528 TGGTACCGGAG-----TCC-----C----- 543
Db 947 TGG-GCCCGGGCTGTTCGCGGCTGGCTGCCCTGAGTACCTTACCTGCAGGACAAC 1005
QY 544 GAG-----AGGCACT-----CAGCTT-----AGCAGG----- 566
Db 1006 GCGCTGAGGCACTGCTGTGATGACACCTTCGCGACCTTGGGCAACCTCACACCTTTC 1065
QY 567 -----TGAAA-----TCCTAC-----AGCGCGTCATC-----GACT--ACATCTC 601
Db 1066 CTGACGACRACCGCATCTCCAGGTGCCCGAGCGCGCTTCCGTGGGCTGCACAGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
Db 1126 GACCGCTCTCTACTGACAGAACCGCTGGCGCCATGTGACCCGCGATGCTTCCGTGAC 1185
QY 618 --TGGCGG-----AGCCAGC----- 630
Db 1186 CTTGCGCGCTCATGACACTCTATCTGTTTGGCAACATCTATCAGCGCTGCCACTGAG 1245
QY 631 -CCCTGGACCCCTGATGCGCC--CC--ACCT-----TC-----CATCC-----A 666
Db 1246 GCGCTGGCGCGCTGCGTGCCTGCGATACCTGAGGCTCAACAGACAACCCCTGGGTGTG 1305
QY 667 GACAGCGAGCTCG---CTCGGNACTTGT-----CATCTCAACAGACAAA 709
Db 1306 GACTCGCGGACGCGCCACTCTGGGCTGGGTGGAAGTTCGGGGCTCTCTCCCGAG 1365
QY 710 AGGAGCTTTTGGC-----ACTGACTCGCGGTGTCTCT--GACACCT-----CCA- 751
Db 1366 GTGCGCTGACGCTCCCGCAACGCGCTGGCT--GGCGGTGACCTCAACGCGCTAGTGCAA 1424
QY 752 -GAAC-GCAGG--TGC-----TGGC-----GCCCGT-----TCTG--C-CTGGGA 784

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C168
CURRENT APPLICATION NUMBER: US/10/140,472
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 381
LENGTH: 2236
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-472-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGGTGTCTGCAGGAAAGCAAAATTC--TGGAAGTTAATGTTTGA-----GTG 54
Db 191 GGCCGGCGGGGGAACCGG-----GCGGATTCCTCGCGCTCAACACCACTGATCCCAT 244
QY 55 ATTTTAAATCTTCTCTGCGG-----GAGAGCGCGCTCTCCCGGTATCAGC 103
Db 245 AAACATTCATCTC--CCGGCGCCCGCTGGAGCGCCCGCCAGTCGCGCCGCC 303
QY 104 GTTCTCTCATTTTGAATCC--GCG--GCTCCGCGGTCTTCGCGCTC---AGACCAAGCCGG 158
Db 304 GC--CCTCGCCCTGTGCGCCCTGCGCGCCCTGCG--CACCGCGCGCCCGAGCCAGCCAG 359
QY 159 AGGAGCGCTGTTGCAATTTAAGCGGGCTGTG-AAAG--CCCAAGGCC--GGCGGGGGCGGG 215
Db 360 AGCCGGCGG--AGCG--GAGCGCGCGGAGCTCTGTCGCGCGCGCGCGCGGGCGGGCGGG 413
QY 216 GCGGAGGCGG--GCCATTTTGAATAAGA---GGC-----GTGCTTCTCA 255
Db 414 GCGGTAGCGCGCGGCC--TGGATCGGACCGCGCGCGCGGAGAGCGCGCGCCGCC 469
QY 256 GCGAGCTCTATAGTGAACCGCGG-----GCGAGCGTGG 292
Db 470 CGAAACGATTTTCAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
QY 293 C-----GC--GTTGCAAGTCA-----CTGTAGC---GGACT-TCCTTTTGTT--T 329
Db 530 CTGGAGGAGCGCGCTGTG--CATGGGTGTGTGTGGCTGAGGCTGCGAGGTGCGACCC 588
QY 330 TCTTTCTCT--TTGG--GCACCTCTGACTCA-----CT--CCCCAG 366
Db 589 CCATGCCCGGCTGCTGCTATGCTACAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 648
QY 367 CA-----TGAAGCGCTGAGCCG--GTGCGCGGTGTACGAGCGGTGTG--CTGCGCTG- 418
Db 649 CAGGCGCTGCG--GCTGTGCGCGTGGGATCCCTGCTGCGAGCGAGCGATCTTCCTGC 706
QY 419 TCGGAA-----CGCA--GTCTG--GCATCGCCCGG--GCCGAG----- 452
Db 707 ACGGCAACCGCATCTCGCATGTGCGAGCTGCGAGCTTCCGTCTGCGCGCAACCTCACA 766
QY 453-----GGAAG-----GGCCCG-----GCAGCTG----- 470
Db 767 TCTGTGCTGCACTCGAATGTGTGGCGCGCGGAAATGATGCGGCTGCTTCACTGCGCTGG 826
QY 471-----AGGAGCGCTG-----AGC-----TTGC-----TGG----- 491
Db 827 CCCTCTCTGAGCAGCTGGACCTCGACCTATGACAGCTCGGTCTGTGGACCTGCGCA 886
QY 492-----AGCAGTGAACCA-----TGCTACTCC-----C-GCCTGGGGGAG 527
Db 887 CATTCCAGGCGCTGGGCGCGCTACACAGCTGACCTGGAGCGCTGCGCGCTGCGAGGAGC 946
QY 528 TGGTACCGCGGAG-----TCC----- 543
Db 947 TGG-GCCCGGGGCTGTTCGCGGCGCTGCTGCGCGCTGCTGCTTCTTCTGCGAGCAAC 1005

Db 1425 TGACCTGAGGGCTGCGCTGTGGCCACCGGCGCTTACATCCATCTGGACCGGCGGCG 1484
QY 785 CCCCG--GGAACCTCT-----CCTCGGGAACCG--GAC--GG- 817
Db 1485 CACCGATGAGGAGCGCTGGGGCTTCCCAAGTGTGCGAGCAGATGCGCTGACAAGGC 1544
QY 818 --CAGGGATGG--GCC-----CCAACCTTG----- 838
Db 1545 CTCAGTACTGAGCTGGAACACAGCTTCGGCAGGCAATGCGCTGAAGGAGCGCTGGCC 1604
QY 839 --CC-----CTGCCACCTGACTTC-----ACCAATTCCTTCC-----T 871
Db 1605 GCCCGGTGACACCGCGCGGCAACCGCTCTGGCCCAACCGGCACATCAATGACTCACCGCT 1664
QY 872 GGAGACTAAACCTGG-----TGCTCA-----GGAGCCGAAG--ACTGT 907
Db 1665 TGGGACTCTGCTGCTGTGAGCCCCCGCTCAGTGCAGTGGCGGCCGAGGCTCCGA 1724
QY 908 G-----AACTTG-----TAGCCTGAAGAGCCAG-----AGCTAGCTC 939
Db 1725 GCCACCGGTTCCCACTCGGCCCTTCCCGGA--GGCAGGCTGTTACGCAAGAAC 1782
QY 940 TG-----GCCA--CCAGCTGGC-----GAGC----- 959
Db 1783 GCGACCGCAGCCACTGCGCTTGGGCGAGCGAGCGAGCGGGGTGGCGGACTGGTGAC 1842
QY 960 -----TCA-----CCTGCTC-----CCACCCCAACCC----- 982
Db 1843 TCAGAGGCTCAGGTGCGCTACCCAGCTCAGCTCAGCTCACCCCTTGGGCGTGGCG 1902
QY 983 CAAGTCTAAGGTCTTTTC-----AGAGCGTG----- 1009
Db 1903 CTGGTGTGTGAGCAGTGTGGGCGCTTGGGCGCTGACCCCGCGGACAGAGGCTGTCTCA 1962
QY 1010 -----GAGGTGTG-----GAAGAGTGGCTGTCTCCAACTATGCCAAGC-----GG 1053
Db 1963 CGAGCAGGTGTGTACATAGGGGT--CT-CTCTCAAGC--CGCAAGCAGCGCGG 2017
QY 1054 CGGCAGA-----GCTGGTCT--TCTGGT--CTCCTTGAAGAAAGTTCGTGTGCGCTG 1102
Db 2018 CGGCGGACCCGTGGGCGGAGCGCGAGGCTCTCTCTGATGAGC--CCTGCGCGCGCG 2075
QY 1103 ATTTATGACCTA--TAATAGATATATAGTTTG-----TACCTTTTTCACAGGAG 1155
Db 2076 CACCCCGATCTCCACCCCATATGTTTACAGGTTGCGGCGAGCGGTTGTTCCAGAACG 2135
QY 1156 GTGACTTTCTGTAAATGCGATGTATA---TTAAACTTTTATAAAAGTT 1203
Db 2136 CCGCTCCACCCAGATCGCGGTATATAGATATGCAATTTATTTTACTT 2186

RESULT 4
US-10-140-472-381
Sequence 381, Application US/10140472
Publication No. US2003013888A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Collin K
APPLICANT: Wood, William

QY 1103 ATTTATGAACCTTA--TAATAGATATATAGGTTTGG-----TACCTTTTATCAGGAAG 1155
Db 2076 CACCCCATCTCCACCCCATCATCTTTACAGGGTTCGGCGCAGCGTTGTTCAGAACG 2135
QY 1156 GTGACTTTCTGTAAACAATGCGATGATA---TAAACTTTTATAAAAAGTT 1203
Db 2136 CGGCTCCACCCAGATCGCGGTATATAGAGATATGATTTTATTACTT 2186

RESULT 5
US-10-141-761-381
; Sequence 381, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-761-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6,2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGGTGTGCGCAGGAAAAAGCAAAATTC--TGGAGTTAATGTTTGA-----GTG 54
Db 191 GGCCGCGGGGAAACCGG-----CGGATTCCTCGCGGTCAAACCACTGATCCCAT 244
QY 55 ATTTTAAATCCTTGTGCGG-----GAGAGCCCGCTCTCTCCCGGTATCAGC 103
Db 245 AAACATTCATCTC--CGGCGCGCGCGCTGCGAGCGCCCGCCAGTCGCGCGCGCGCC 303
QY 104 GCTTCTCATTTCTTTGAATCC--GCG-GTTCGCGGTCTTCGGCGTC---AGACCAAGCCGG 158
Db 304 GC--CTTCGCTCTGTGCGCTCTGCGCGCTCG--CAACCGCGCGCGGAGCCAGCCAG 359
QY 159 AGGAAGCCTGTGTGCAATTTAAGCGGCTGTG-AAAG--CCAGAGGCC--GGCGGGGGCGG 215
Db 360 AGCGGGCGG--AGCG--GAGCGCGCGAGCTCTCTCCCGCGCGCGCGCGCGCGCGG 413
QY 216 GCGGAGCGG--GCAATTTGAATAAGA--GSC-----GTGCTTCCA 255
Db 414 GCGGTAGCGCGCGGCG--TGATGCGACCCGCGCGGGGAGACGGGCGCGCGCC 469
QY 256 GCGAGGCTCTATAAGTGAACCGCGG--G-----GCGAGCGTGG 292
Db 470 CGAAACGACTTTTCAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 529
QY 293 C-----GC--GTTGAGGTCA-----CTGTAGC-----GGACT--TCTTTTGGTT--T 329

QY 544 GAG---AGGCACT-----CAGCTT---AGCCAGG-----566
Db 1006 GCGCTGAGGACCTGCTGTATGACACTTCGCGGACTGGGCAACCTCAGACCTCTTC 1065
QY 567 -----TGAAAA-----TCTAC-----AGCGGTATC-----GACT--ACATCTC 501
Db 1066 CTGACGGCAACGCACTCTCAGCGTCCCGAGCGCGCTTCGCTGGGTGACAGCCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTC-----617
Db 1126 GACGCTCTCTACTGCAACCGCGTGGCCATGTGACCCGGATGCTTCCGTGAC 1185
QY 618 --TGCGCG-----AGCCAGC-----630
Db 1186 CTTGGCGCTCATGACACTCTATCTTTGCCAAACAATATCAGCGCTGCCCACTGAG 1245
QY 631 -CCCTGACCCCTGATGGCC--CC--ACCT--TC-----CCATCC-----A 666
Db 1246 GCGCTGGCCCCCTCGCTGCGCTCTGAGTCTGAGGCTCAAGACACCTCGGTGTGT 1305
QY 667 GACAGCGGAGCTCG-----CTCGGAACCTGT-----CATCTCAACGACAA 709
Db 1306 GACTGCGGGCAGCCCACTCTGGCTGGTGCAGAAATTCGCGGCTCTCTCTCGAG 1365
QY 710 AGGAGCTTTTGC-----ACTGACTGGCGGTCTCT--GACACCT-----CCA- 751
Db 1366 CTGCGCTCAGCCTCCCGCAACGCTGCT--GGCGGTGACCTCAACGCTTACTGCGCA 1424
QY 752 -GAAC-GCAGG--TGC-----TGCG-----GCGCT-----TCTG--C--CTGGGA 784
Db 1425 TGACTGAGGCTGCGCTGTGGCCACCGGCTTACCATCCATCTGGACCGCAGGCG 1484
QY 785 CCGCG-----GGAACCTCT-----CTGCGGAACCG--GAC--GG--817
Db 1485 CACCGATGAGGCGCTGGGGCTTCCCAAGTGTGCGCAGCAGATCCCGTGCACAGGC 1544
QY 818 --CAGGGATG--GCC-----CAACTTCG-----838
Db 1545 CTCAGTATGAGCTGGAAGCAGCTTCGGCAGGCAATGCGCTGAAGGACGCGTGC 1604
QY 839 --CC-----CTGCCACCTGATTC-----ACCAATCCCTTC-----T 871
Db 1605 GCGCGGTGACACCGCGCGGCAACGCGCTGTGGCCACGCGCACATCAATGACTCACCT 1664
QY 872 GGAGACTAACTTG-----TGCTCA-----GGAGCGAAGG--ACTGT 907
Db 1665 TGGGACTTGCCTGCTGTGCTGAGCCCGCTCACTGAGCGCGCGCGGAGCTCGA 1724
QY 908 G-----AACTTG-----TAGCTGAAGAGCAG-----AGTAGCTC 939
Db 1725 GCCACAGGGTTCGCCACTCGGGCCCTCGCGGA--GGCCAGGCTGTTCACGCAAGAC 1782
QY 940 TG-----GCCA--CCAGCTGGC-----GAGC-----959
Db 1783 CGACCGCAGCCTGCGCTGTGGCCAGGAGCAGCGGGGTGGGAGCTGGTAC 1842
QY 960 -----TCA-----CCCTGCTC-----CCACCCCAACCC-----982
Db 1843 TCAGAGGCTCAGGTGCGCTACCCAGCTCAGCTGACGCTCACCCCGCTGGCGTGGCG 1902
QY 983 CAACTTCAAGCTCTTTT-----AGAGCGTG-----1009
Db 1903 CTGGTCTGTGACAGTGTCTGGGCGCTGACCCCGAGCGGACAAAGAGCTGTCTCA 1962
QY 1010 -----GAGGTG-----GAAGGAGTGTCTCTCCAAACTATGCCAAGC-----GG 1053
Db 1963 GAGCGAGTGTGTACATACGGGT--CT--CTCTCCAGC--CGCCCAAGCCAGCGGG 2017
QY 1054 CGGACAG-----GCTGGTCT--TCTGGT--CTCTTGGAGAAAGTTCTGTGGCTG 1102
Db 2018 CGGCCAGCCGTGGGCGAGCGCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2075

Db 530 CTGAGGGAGCCGCTGCTGG-CATGGTCTGTGCTGCAGGCTGGAGTGGCAGCC 588
 QY 330 TCTTTCTCT-----TTGGG--GCACCTCTGGAATCA-----CT--CCCGAG 366
 Db 589 CCAATGCCAGGTGCTGCTGATCTCAATGAGCCCAAGTGACGACAGCTGCCCCAG 648
 QY 367 CA-----TGAAGCGCTGAGCCCG-GTGCGCGGCTGTAGAGGCGGTGTG-CTGCTG- 418
 Db 649 CAGGGCTGCAG--GCTGTGCCGTGGGATCCCTGCTGCCAGCCAGCGCATCTTCTGTC 706
 QY 419 TCGGAA-----CGCA--GTCTG--GCCATCGCCCGG--GCCGAG----- 452
 Db 707 ACAGCAACCGCATCTGCGATGTCAGCTGCGAGCTTCGCTGCGCGCAACCTCACCA 766
 QY 453 -----GGAAG-----GGCCCG-----GCAGTG----- 470
 Db 767 TCCTGTGGTGCACATCGAATGTGCTGCGCCGGAATTGATCGGCTGCTTCACTGGCGTGG 826
 QY 471 -----AGGAGCGCTG-----AGC--TTGC-----TGG----- 491
 Db 827 CCCTCTGGAGAGCTGGACCTCAGCGAATGCAACAGCTCCGGTCTGTGGACCTTGCCA 886
 QY 492 -----ACGACATGAACAC-----TGCTATCC-----C-GCCTGGGGGAAAC 527
 Db 887 CATTCACGGCTGGGCGGCTACACACAGCTGCACCTGGACCGCTGCGGCTGCAGGAGC 946
 QY 528 TGGTACCGGAG--TCC-----TCC-----C 543
 Db 947 TGG-GCCCGGGGTGTTCCGCGCTGGCTGGCTGCGCTGAGTCTTACCTGCGAGCAAC 1005
 QY 544 GAG-----AGGCACT-----CAGCTT-----AGCCAG----- 566
 Db 1006 GCGCTGCAGCACTGCTGTGATGACACCTTCGCGAGCTGGGCAACCTCACACCTCTTC 1065
 QY 567 -----TGGAA-----TCTTAC-----AGCGGTGATC-----GACT--ACATTC 601
 Db 1066 CTGCACGGCAACCGCATCTCAGCGTGCACGCGCGCTTCCGTGGCTGCAACCTTC 1125
 QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
 Db 1126 GACGCTCTCTACTGACACAGACCGGCTGGCCCATGTGCAACCGCATGCTTCCGTGAC 1185
 QY 618 --TGGCGC-----AGCCAG----- 630
 Db 1186 CTTGGCGCTCATGACACTTATCTGTTTGGCAACAATATCAGCGTGGCCCATGAG 1245
 QY 631 -CCCTGACCCCTGATGGCC--CC--ACCT--TC-----CCATCC-----A 666
 Db 1246 GCGCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
 QY 667 GACAGCCGAGTGG-----CTCGGAACTTGT-----CATCTCCAGCAAA 709
 Db 1306 GACTGCCGGGCAACCCCACTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
 QY 710 AGGAGCTTTGCC-----ACTGCTGCGCGGTGCT--GACACCT-----CCA- 751
 Db 1366 GTGCCCTGAGCCTCCGCAAGCCTGGCT--GGCCGTAACCTAAGCCTAGCTGCGCA 1424
 QY 752 -GAAC-GCAGG--TGC-----TGGC-----GCCCGT-----TCTG--C-CTGGGA 784
 Db 1425 TGACCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1484
 QY 785 CCCC-----GGAACCTCT-----CCTGCGGAGCGG--GAC--GG- 817
 Db 1485 CACCGATGAGAGCGCTGGGGCTTCCCAAGTGTGCTGCGCAGCATGCTCCCTGACAGGC 1544
 QY 818 --CAGGATGG-GCC-----CCAACCTTCG----- 838
 Db 1545 CTCAGTACTGAGCTGGAAGACCACTTGGCAGGCAATGCGCTGAAGGAGCGGTGCC 1604
 QY 839 --CC-----CTGCCCACTGACTTC-----ACCAATCCCTTC-----T 871

Db 1605 GCCCGTGCAGAGCCCGCGGGCAACGGCTCTGCGCCACGCGCATCATCACTGACTCACCTT 1664
 QY 872 GGAGACTAAACCTGG-----TGCTCA-----GGAGCGAAGG-ATGT 907
 Db 1665 TGGGACTCTGCTGGCTCTCTGAGCCCGCTCTGAGTGGCGCCGAGGGCTCCGA 1724
 QY 908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGTAGCTC 939
 Db 1725 GCAACACAGGTTCCCACTCGGGCCCTCGCGGA--GGCAGGCTGTTTCAAGCAAGAAC 1782
 QY 940 TG-----GCCA--CCAGCTGGG-----GAGC----- 959
 Db 1783 GCAACCGGACCACTGCTGCTGCGGCGAGGAGCGGGGTGGCGGACTGGTGCAC 1842
 QY 960 -----TCA-----CCCTGCTC-----CCACCCCACTCC----- 982
 Db 1843 TCAGAGGCTCAGGTGGCTTACCCAGCTCAGCTGAGCTCACCCCTGGGCTGGCG 1902
 QY 983 CAAGTCTTAAGGTCTTTTC-----AGAGCGTG----- 1009
 Db 1903 CTGTGCTGTGAGCAGTGTGTTGGCCCTGTGACCCCGGAGCGGACACAGAGCGTGTCA 1962
 QY 1010 -----GAGTGTG-----GAAGGAGTGGCTCTCCAACTATGCAAGG-----GG 1053
 Db 1963 CGAGCAGGTGTGTACATACGGGT--CT-CTCTCCAGC--CGCAGCGCAGCGGG 2017
 QY 1054 CGGCGA-----GCTGTCT--TCTGTG--CTCCTGAGAGAAAGTCTGTGTGCGCTG 1102
 Db 2018 CGGCGACCCGTGGGCGAGCGCAGCGAGTCTCTCTGTGAGC--CCTGCGCGCGC 2075
 QY 1103 ATTTATGACTCTA--TAATAGATATATAGTGTG-----TACCTTTTTCAGGAAG 1155
 Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTGCGGCGAGCTTGTTCAGAACG 2135
 QY 1156 GTGACTTCTCTAAACATGCGATGTATA--TTAAACTTTTATAAAGTT 1203
 Db 2136 CGGCTCCACCCAGATCGCGGTATATAGATATGATTTTATTACTT 2186

RESULT 6

US-10-142-885-381
 ; Sequence 381, Application US/10142885
 ; Publication No. US20030157604A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tamas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P33301C248
 ; CURRENT APPLICATION NUMBER: US/10/142,885
 ; CURRENT FILING DATE: 2002-05-10
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 381
 ; LENGTH: 2236
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 ; US-10-142-885-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
 Best Local Similarity 39.8%; Pred. No. 6.2e-171;
 Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGTGCTGCCAGGAAAGCAATTC--TGGAAGTTAAATGTTTCA---GTG 54
 Db 191 GCGCGCGGGGAAACCGG-----GCGGATCTCCGCGGTCAACCACTGATCCATA 244

QY 55 ATTTTAAATPCCTTGTGGCG-----GAGAGGCCCGCTTCCCGGATATCAGC 103
 Db 245 AAACATTCATCTTC--CGGCGCGCGCGGTGAGCGCGCGCGCAGTCCGCGCGCGCC 303

QY 104 GCTTCCTCATCTTTGAATCC--GG--GCTCCGGGCTTCGGGCTC--AGACCAAGCGG 158
 Db 304 GC--CTTCGCGGCTGTCGCGCTCGCGGCTCGG--CAACCGCGCGCGAGCCAGCCAG 359

QY 159 AGGAAGCTGTTTGAATTTAAAGCGGCTGTG--AACG--GCCAGGCTC--GGCGGGGCGG 215
 Db 360 AGCGGCGCG--AGCG-----GAGCGCGCGAGCTCTGTCGCGCGCGCGCGCGCG 413

QY 216 GCGGAGCGG--GCCATTTGAATAAGA--GGC-----GTGCTTTCCA 255
 Db 414 GCGGTAGCGCGCGCGC-----TGATCGGACCGCGCGCGGGGAGACGGGCGCGCC 469

QY 256 GCGAGGCTATTAAGTGACCGCGCG-----GCGAGCGTGGC 292
 Db 470 CGAAGACATTCAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529

QY 293 C-----GG--GTTCCAGGTCA-----CTGTAGC--GGACT--TCTTTTGGTT--T 329
 Db 530 CTGAGGAGCGCGGCTCTG--CATGGGTCTGTGGCTGCGAGGCTTGGAGGTGGCAGCC 588

QY 330 TCTTTCT-----TTGGG--GCACCTCTGGACTCA-----CT--CCCCAG 366
 Db 589 CCATGCCAGGTGCTCGTATGCTCAATGAGCCGAGGTGACGACAGTGCCTCCCGAG 648

QY 367 CA-----TGAGGCGCTGAGCCCG--GTGCGGCTGCTACAGCGGTGTG--CTGCTG-- 418
 Db 649 CAGGCGCTGCAG--GCTGTGCGCGTGGGCATCCCTGTGCGCAGCGCATCTTCTG 706

QY 419 TCGGAA-----CGCA--GPTCTG--GCCATCGCCGCGG--GCCGAG----- 452
 Db 707 ACGGACCGCATCTCGCATGTGCGAGTGCAGTGCAGCTTCGCTGCGCGCAACTCAGCA 766

QY 453 -----GGAAG-----GCCCCG-----GCAGTG----- 470
 Db 767 TCCTGTGGCTGCACTCGAATGTGCTGGCCGGAATTGATGGGCTTCACTTGGCGCTG 826

QY 471 -----AGGCGCGCTG-----AGC-----TTGC-----TGG----- 491
 Db 827 CCCTCTGGAGCGCTGGAACCTAGCGATATGACAGCTCCGGTCTGTGGAACCTTGCCA 886

QY 492 -----ACGACATGAACAC-----TGCTACTTC-----C--GCTTCGGGAAC 527
 Db 887 CATTCACGGCTTGGCGCGCTACACAGCTGCACCTGGAGCGGTGGCGGCTGAGGAGC 946

QY 528 TGGTACCGGAG-----TCC-----C----- 543
 Db 947 TGG--GCCCGGGCTGTTCCCGGCGCTGGCTGCCGTGAGTACCTACCTGCGAGCAAC 1005

QY 544 GAG-----AGGCACT-----CAGTT-----AGCCAGG----- 566
 Db 1006 GCGTGCAGGCACTGCTGTATGACACTTCCGCGACCTGGGCAACCTCACACCTCTTC 1065

QY 567 -----TGAAA-----TCCTAC-----AGCGGTATC-----GACT--ACATCTC 601
 Db 1066 CTGACGCGCAACCGCATCTCCAGCGTGGCCGAGCGCGCTTCCGTGGGCTGACAGCCTC 1125

QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
 Db 1126 GACCGTCTCTACTGACACCAAGACCGCGTGGCGGCTGACCGCGCATGCTTCCGTGAC 1185

RESULT 7

US-10-158-790-381

; Sequence 381, Application US/10158790

; Publication No. US20030180879A1

QY 618 --TGGCGG-----AGCCAGC----- 630
 Db 1186 CTTGGCGGCTCATGACACATCTATCTGTTTGGCAACAATCTATCAGCGTGCACCTGAG 1245

QY 631 -CCCTGACCCCTGTATGGGCC--CC--ACCT--TC-----CCATCC-----A 666
 Db 1246 GCGCTGCGCGCGCTGCTGCGCTGAGTACCTGAGGCTCAACGACAAACCCCTGGGTGT 1305

QY 667 GACAGCGAGTCTG-----CTCCGGAATTTG-----CATCTCCAGCAAA 709
 Db 1306 GACTCGCGGACGCCCACTCTGGGCTGTGCTGAGAAAGTTCGCGGCTCTCTCCGAG 1365

QY 710 AGGAGCTTTTGGC-----ACTGACTCGGCGGTGTCT--GACACT-----CCA- 751
 Db 1366 GTGCCCTGCACTCTCCCGCAACGCTGGCT--GGCGGTGACTCAAAAGCGCTAGTGCAA 1424

QY 752 -GAAC--GCAGG--TGC-----TGGC-----GCGCGT-----CTG--C--CTGGGA 784
 Db 1425 TGACTCGAGGCTGCGCTGTGGCCACCGGCGCTTACCATCCCATCTGAGACCGGAGGC 1484

QY 785 CCGCG-----GGAACCTCT-----CCTGCCGGAAGCG--GAC--GG-- 817
 Db 1485 CACCGATGAGGAGCGGCTGGGCTTCCAAAGTGTGCCAGCAGATGCCCTGACAAGGC 1544

QY 818 -CAGGATGG--GCC-----CCAACTTCG----- 838
 Db 1545 CTCAGTACTGAGGCTGGAAGACCACTTCGGCAGGCAATGCGTGAAGGACGCGTGCC 1604

QY 839 -CC-----CTGCCCACTTGACTTC-----ACCAAACTCCCTTC-----T 871
 Db 1605 GCGCGGTGACAGCGCGCGGCAAGGCTCTGGGCCACGCGCATCAATGACTCACCTT 1664

QY 872 GGAGACTTAACTCTG-----TGCTCA-----GGAGCGAAGG--ACTGT 907
 Db 1665 TGGACTCTGCTGGCTGTGCTGAGCCCGCTCACTGCAAGTGGCGGCGGAGGCTCCGA 1724

QY 908 G-----AATTG-----TAGCTCAAGAGCCAG-----AGTAGTCTC 939
 Db 1725 GCGACAGGTTCCCACTCGGCGCTCGCCGGA--GGCCAGGCTGTTCAGCAAGAAC 1782

QY 940 TG-----GCCA--CCAGCTGGG-----GACG----- 959
 Db 1783 GCGACCGCAGCACTGCGCTGTGGCGCAGGCGGCGGCTGGCGAGCTGGTGAC 1842

QY 960 -----TCA-----CCCTGCTC-----CCACCCCAACCC----- 982
 Db 1843 TCAGAGGCTCAGGTGCGCTTACCAGGCTCACTGAGCTCACCCCTGGGCTGGCG 1902

QY 983 CAAGTTCTAAGTCTTTTC-----TGCGGCTGTGACCCCGGAGACAAAGAGCGTGTCA 1009
 Db 1903 CTGGTGTGTGACAGTCTTGGGCGCTGTGACCCCGGAGCGGAGCGTGTCA 1962

QY 1010 -----GAGTGTG-----GAGGAGTGGCTGCTCCAACTATGCCAAGC-----GG 1053
 Db 1963 GCAGCGAGGTGTGTATACAGGCT--CT--CTCTCCAGC--CGCCAGCCAGCGGG 2017

QY 1054 CGGCGAGA-----GCTGGTCT--TCTGGT--CTCCTTGGAGAAAGTTCGTTCGCTG 1102
 Db 2018 CGGCGAGCCCGTGGGCGAGGCGCAGGCTCTCTCCCTGATGGAGC--CCTGCGCGCGC 2075

QY 1103 ATTTATGACTCTA--TAATAGATATATAGTTTG-----TACCTTTTACAGGAAG 1155
 Db 2076 CACCGCCATCTCCACCCCATCATGTTTACAGGTTTGGCGGCGAGGTTTTCAGAACG 2135

QY 1156 GTGACTTTCTGTAAACAATGCGATGTATA---TTAACTTTTATAAAAGTT 1203
 Db 2136 CCGCTCCCAACCGCATCGGATATAGAGATATGCAATTTTACTTT 2186

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
 APPLICANT: Beresini, Maureen
 APPLICANT: DeForge, Laura
 APPLICANT: Desnoyers, Luc
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Sherwood, Steven
 APPLICANT: Smith, Victoria
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tunas, Daniel
 APPLICANT: Watanabe, Colin K
 APPLICANT: Wood, William
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3330R1C448
 CURRENT APPLICATION NUMBER: US/10/158,790
 CURRENT FILING DATE: 2002-05-30
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 550
 SEQ ID NO 381
 LENGTH: 2236
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-10-158-790-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
 Best Local Similarity 39.8%; Pred. No. 6.2e-171;
 Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGGTCTCCAGGAAGAAGCAATTC-TGGAGTTAATGTTTGA---GTG 54
 DB 191 GGCGGGGGGAAACCGG-----GCGGATTCCTGGCGGTCAACACCTGATCCATA 244
 QY 55 ATTTTAAATCCTTGCTGGCG-----GAGAGCGCGCTCTCCCGGATATCAGC 103
 DB 245 AAACATTATCTC-CGGGCGCGCGCTGCGAGCGCCCGCGAGTCCGCGCGCGCC 303
 QY 104 GTTCTCTATCTTTGAATCC-GCG-GCTCGCGGTCTTGGGTC---AGACACCGCG 158
 DB 304 GC--CCTCGCGCTGCGCGCTGCGCGCTGCG--CACCGCGCGCGAGCCAGCCAG 359
 QY 159 AGGAAGCTGTTCGAATTTAAGCGGCTGTG-AAACG-CCACAGGCG-CCGCGGGCGGG 215
 DB 360 AGCGGGCGG--AGCG---GAGCGCGCGAGCTCTGCGCGCGCGCGCGCGCGCG 413
 QY 216 GCGAGGCGG---GCCATTTGAATAAGA---GGC-----GTGCTTCCA 255
 DB 414 GCGGTAGCGGGCGGCG---TGATGCGGACCGCGCGCGCGGAGACGGCGCGCGCC 469
 QY 256 GCGAGGCTATAGTGAACCGCGCG---GCGAGCGGTGCG 292
 DB 470 CGAAAGCACTTCTAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
 QY 293 C---GC--GTTCAGGTCA-----CTGTAGC---GACT-TCCTTTGGTT--T 329
 DB 530 CTGAGGAGCGCGGTGCTGCG-CATGGTGTGTGTGTGCGAGGCTGCGAGGTGCGAGCC 588
 QY 330 TCTTCTCT---TTGGG---GCACCTCTGACTCA-----CT--CCCCAG 366
 DB 589 CATGCCAGGTGCTGCTGTATGTAATGAGCCAGGTGAGCAGCAAGCTGCCCGCAG 648
 QY 367 CA-----TGAAGGGCGTGAAGCCG-GTGGCGGCTGTGAGAGCGGTGTG-CTGCTG- 418
 DB 649 CAGGCGCTGCAG--GCTGTGCGCGTGGGCACTCCCTGCTGCGAGCCAGCGCATCTTCTG 706
 QY 419 TCGGAA-----CGCA---GTCTG--GCATCGCCCGG---GCCGAG----- 452

DB 707 ACGCAACCGCATCTGCAATGTGCGAGCTGCGAGCTTCGTGCTCGCGCAACCTCACCA 766
 QY 453 -----GGAAG-----GCGCG-----GCAGCTG----- 470
 DB 767 TCCTGTGGCTGCACTCGAATGTGCTGCGCGCAATGATGCGGCTGCTTCACTGGCTGG 826
 QY 471 -----AGGAGCGCGCTG-----AGC-----TTGC-----TGG----- 491
 DB 827 CCCTCTGGAGCAGCTGACCTGACGATATGACAGCTCCGCTCTGTGGACCTGCCA 886
 QY 492 -----ACGACATGAACAC-----TGCTACTCC-----C-GCTGGGGAAC 527
 DB 887 CATTCACGCGCTGGCGCGCTACACAGCTGCACTTGGACCGCTGCGGCTGCGAGGC 946
 QY 528 TGGTACCGGAG---TCC-----TCC-----C 543
 DB 947 TGG-GCGCGGCTGTTCCGCGGCTGGCTGCTGCGAGTACCTTACCTGCGAGCAAC 1005
 QY 544 GAG-----AGGCACT-----CAGCTT-----AGCAGG----- 566
 DB 1006 GCGCTGCGAGCAGCTGCTGATGACACCTTCCGCGACCTTGGGCAACCTCACACCTTTC 1065
 QY 567 -----TGGAAA-----TCCTAC-----AGCGGTGATC-----GACT--ACATTTCT 601
 DB 1086 CTGACGCAACCGATCTCCAGGTGCGCGAGCGGCTTCCGTGGGCTGACAGCTC 1125
 QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
 DB 1126 GACGCTCTCTACTGACAGAACCGGTGCGCGCTGCAACCGCATGCTTCCGTGAC 1185
 QY 618 -TGGCGG-----AGCCAGC----- 630
 DB 1186 CTGCGCGCTGATGACACTTATCTGTTTCCAACTATCATGAGGCTGCCACTGAG 1245
 QY 631 -CCCTGGACCCCTGATGGCC--CC-ACCT--TC-----CCATCC-----A 666
 DB 1246 GCGCTGGCGCGCTGCGTGGCTGAGTACCTGAGGCTCAACGACACCCCTGGGTGTGT 1305
 QY 667 GACAGCGAGCTCG---CTCGGAACCTGT-----CATCTCAACGACAAA 709
 DB 1306 GACTCCGCGGACGCGCACTCTGCGGCTGGCTGAGAGTTCCGCGGCTCTCTCCGAG 1365
 QY 710 AGGAGCTTTTGGC-----ACTGACTCGCGGCTGTCT--GACACCT-----CCA- 751
 DB 1366 GTGCGCTGACGCTCCCGCAACGCTGCT--GGCGTGACCTCAACGCTAGTGCCAA 1424
 QY 752 -GAAC-GCAGG--TGC-----TGGC-----GCGCGT-----TCTG--C--CTGGA 784
 DB 1425 TGACCTGAGGCTGCTGTGCGCACCGGCTTACCATCCATCTGACCGGAGGCG 1484
 QY 785 CCGCG---GGAACCTCT-----CCTGCGGGAAGCGG--GAC--GG- 817
 DB 1485 CACCAATGAGAGCGCTGGGCTTCCCAAGTGTGCTGCCAGCAGATGCGGTGACAAG 1544
 QY 818 --CAGGATGG-GCC-----CMACTTCG----- 838
 DB 1545 CTGAGTCTGAGCGCTGGAAGACCACTTCCGCGAGGCAATCGCTGAAGGACCGGTGCC 1604
 QY 839 --CC-----CTGCGGCTTGTGCTTC-----ACCAATCCCTTCG-----T 871
 DB 1605 GCGCGGTGACCGCGCGCGGCAAGGCTCTGGCGGCAAGCATCAATGACTCACCTT 1664
 QY 872 GGAGACTAAACCTGG-----TGCTCA-----GGAGCGAGG-ACTGT 907
 DB 1665 TGGGACTCTGCTGCTGTGAGGCGCGCTCACTGAGTGGCGCGCGGCTCCGA 1724
 QY 908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGCTAGTCTC 939
 DB 1725 GGCACAGGTTCCCGCACTCTGGGCGCTGCGCGGA--GGCCAGGCTGTTCAGCAGAAC 1782
 QY 940 TG-----GCCA---CCAGCTGGGC-----GACG----- 959
 DB 1783 CGCACCGGAGCACTGCGCTCTGGCGGAGGCGGCGGCTGGCGGAGTGGTGTGAC 1842

QY 960 -----TCA-----CCCTGCTC-----CCACCCACCCCC-----982
Db 1843 TCAGAGGCTCAGGTGCTTACCCAGCTCAGCTGACGCTCACCCTGCGCTTGGCG 1902
QY 983 CAAGTCTAAGGTCTTTTC-----AGAGCGTG-----1009
Db 1903 CTGGTGTGTGACAGTGTCTTGGCCCTGTGACCCCGGACAGGAGCGTGTCTCA 1962
QY 1010 -----GAGGTGTG-----GAAGAGTGTGCTTCTCCAAACTATGCCAAGGC-----GG 1053
Db 1963 CGAGCCAGGTGTGTACATACGGGT--CT-CTCTCCACGC--CGCCAAAGCCAGCGGG 2017
QY 1054 CGGCAGA-----GCTGTGCT--TCTGCT-CTCCTTGAGAAAGTCTTGTGCTG 1102
Db 2018 CGGCCGACCGCTGGGGCAGGCGCAGGTCTCTCCCTGTGAGC--CCTGCGGCCCG 2075
QY 1103 ATTATGACCTA--TAATAGATATATAGTTTG-----TACCTTTTATCAGGAAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTTGGCGGACGCTTGTTCAGAACG 2135
QY 1156 GTGACTTTCTGTAACTATGCGATGATA--TTAACTTTTATAAAGTT 1203
Db 2136 CGCCTCCACCCAGATCGGTATATAGATATGCAATTTTATTACTT 2186

RESULT 8

US-10-137-871-381
; Sequence 381, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C153
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-381

Query Match 47.6%; Score 3593; DB 15; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;
QY 1 GATCTGGGGTGTGTCGAGGAAAAAGCAATTC--TGAAGTTAATGGTTTTGA-----GTG 54
Db 191 GGCCGGCGGGGAACCG-----GGGATTCTCTCGCGGTCAAAACCACTGATCCCATTA 244
QY 55 ATTTTAAATCTTGTGCG-----GAGAGGCCGCTCTCCCGGTATCAGC 103
Db 245 AAACATTCATCTC--CGGGCGCCCGGGTGTGAGCGCCCGCCAGTCCGCGCCGCC 303

QY 104 GCTTCTCATTTCTTTGAATCC--GCG--GCTCCGCGGTCTTCCGCGTC--AGACAGCGCG 158
Db 304 GC--CCTCGCCTGTGCGCCCTGCGCCCTGCG--CACCCGCGCGCGAGCCAGCCAG 359
QY 159 AGGAAGCCTGTTTGAATTTAAGCGGGGTGTG--AACG--CCAGAGGCC--GGCGGGGGCGGG 215
Db 360 AGCCGGCGG--AGCG--GAGCGCGCGAGCCTGTCCTCCGCGCGCGCGCGCGCGCG 413
QY 216 GCGGAGGGG--GCCATTTTGAATAAGA--GGC-----GTGCTTTCCA 255
Db 414 GCGGTAGCGCGCGGCC--TGGATGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 469
QY 256 GCGAGGCTCTATAAGTGAACCGCGCG-----GCGAGCGTGC 292
Db 470 CGAAGAGCTTTCAGTCTCCCGAGCGCGCGCCGCCAACCCCTACGATGAAGAGGCGCTCG 529
QY 293 C-----GC--GTTGAGGTCA-----CTGTAGC-----GGACT--TCITTTGGTT--T 329
Db 530 CTGAGGAGCGCGCTGCTG--CATGGGTGTGTGGCTGCGAGGCTTGCAGGTGCGACGCC 588
QY 330 TCTTTCTCT--TTGGG--GCACCTCTGAGTCA-----CT--CCCCAG 366
Db 589 CATGCGCCAGGTGCTGCTGTATGTCTCAATGAGCCCAAGGTGACACAAGCTGCCCGCAG 648
QY 367 CA-----TGAAGCGCTGAGCGCG--GTGCGGGGTGTACGAGGGGTGTG--CTGCTG-- 418
Db 649 CAGGGCTGCG--GCTGTGCCGTGGGCATCCCTGCTGCCAGCGCATCTTCTTCG 706
QY 419 TCGGAA-----CGCA--GTCTG--GCCATCGCCCGG--GCCGAG-----452
Db 707 ACGGCAACCGCATCTGCAATGTGCTGCGCGGAAATGATGCGGCTGCTTCACTGCGCTGG 766
QY 453 -----GMAAG-----GGGCGG-----GCAGCTG-----470
Db 767 TCTGTGGCTGCACTCGAATGTGTGCGCGGAAATGATGCGGCTGCTTCACTGCGCTGG 826
QY 471 -----AGGAGCGGTG-----AGC-----TTGG-----TGG-----491
Db 827 CCTCTCGAGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 886
QY 492 -----ACGACATGAACCA-----TGCTACTC-----C-GCTTGGGGAGAAC 527
Db 887 CATTCACGCGCTGGCGCGCTTACACACGCTGACGCTGAGCGCTGCGCGCTGCGAGC 946
QY 528 TGGTACCGGAG-----TCC-----TGGTACTC-----C 543
Db 947 TGG--GCCCGGGGTGTTCGCGCGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1005
QY 544 GAG-----AGGCACT-----CAGCTT-----AGCCAG-----566
Db 1006 GCGTGCAGGCTGCTGCTGATGACACCTTCCGCGACCTGGGCAACCTCACACCTCTTC 1065
QY 567 -----TGGAAA-----TCCTAC-----AGCGGTCTATC-----GACT--ACATTCTC 601
Db 1066 CTGACGCGCAACCGCATCTCCAGCGTGGCGGCGCGCTTCCGTGGGTGCGACGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC-----617
Db 1126 GACGCTCTCTACTGTGACCAAGACCGGTGGCGCATGTGACCCGCGCATGCTTCCGTGAC 1185
QY 618 --TGGCGG-----AGCAGC-----630
Db 1186 CTGGCGCCTCATGACACTTATCTGTTTGGCAACAATCTATCAGCGTGGCCACTGAG 1245
QY 631 -CCTTGAGCCCTGATGGGCC--CC--ACCT-----TC-----CCATCC-----A 666
Db 1246 GCGCTGGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1305
QY 667 GACAGCGAGCTCG-----CTCCGGAATTTG-----CATCTCAACGACAAA 709
Db 1306 GACTGCGGCGCAGCCCACTCTGGGCGTGTGCTGAGAAAGTTCGCGGCTCTCTCCGAG 1365
QY 710 AGGAGCTTTTGGC-----ACTGACTCGCGCGTGTCT--GACACCT-----CCA- 751

Db 1366 GTGCCCTGCAGCTCCCGCAAGCCTGGCT-GGCCGTGACCTCAAGCGCTAGCTGCCAA 1424
Qy 752 -GAAC-GCAGG--TGC-----TGCC-----GCCCGT-----TCTG---C--CTGGGA 784
Db 1425 TGACCTGCAGGCTGCGCTGTGGCCACCGGCTTACCATCCATCTGACCGCGCAGGGC 1484
Qy 785 CCCC-----GGAACCTCT-----CCTGCCGGAAGCG--GAC--GG--817
Db 1485 CACGATGAGAGCGCTGGGGCTTCCCAAGTGTGCCAGCAGATGCCCTGACAAGC 1544
Qy 818 --CAGGGATG-GGC-----CCAAGTTG-----T 871
Db 1545 CTCAGTACTGAGCTGGAAGACCACTGCGCAGGCAATGCGTGAAGGAGCGCTGCC 1604
Qy 839 --CC-----CTGCCCATGACCTC-----ACCAATCCCTTCC-----T 871
Db 1605 GCCCGGTGACAGCCCGCGGCAACGGCTGTGGCCACGCGACATCAATGACTCACCTT 1664
Qy 872 GGAGCTAAACCTGG-----TGCTCA-----GGAGCGAAGG-ACGTG 907
Db 1665 TGGACTCTGCTGGCTCTGCTGAGCCCGCTCACTGCACTGCGCGCGCGGCTCCGA 1724
Qy 908 G-----AACTTG-----TAGCTGAAGAGCGAG-----ACGTAGCTC 939
Db 1725 GCCACAGGGTTCCCGCACTCGGGCCCTCGCCGA--GGCAGGCTGTTACGCAAGAAC 1782
Qy 940 TG-----GCCA--CCAGCTGGC-----GAGC-----959
Db 1783 CGACCCGACGCACTGCGCTGTGGCCAGCAGCGAGCGGGGTGCGGGACTGGTAC 1842
Qy 960 -----TCA-----CCTGCTC-----CCACCCACCCC-----982
Db 1843 TCAGAGGCTCAGGTGCGCTTACCAGGCTCAGCTGCACTCAGCTCACCCCTCGGGCTGGC 1902
Qy 983 CAAGTTCTAAGCTTTTC-----AGAGGTG-----1009
Db 1903 CTGCTGTGTGGCAGTCTTTGGGCTGTGACCCCCAGCGGACAGAGGTGTCTCA 1962
Qy 1010 -----GAGGTGTG-----GAAGGAGTGTGCTCTCCAACTATGCCAAGC-----GG 1053
Db 1963 GCAGCCAGTGTGTACATACGGGT--CT-CTCTCCAGC--CGCCAGCCAGCGCGG 2017
Qy 1054 CGGCAGA-----GCTGGTCT--TCTGGT-CTCCTTGGAGAAAGTTCTGTGGCCTG 1102
Db 2018 CGCGCAGCCGCTGGGGCAGCGCCAGGCTCCTCCCTGATGAGC--CCTGCCGCCCGC 2075
Qy 1103 ATTTATGAACCTTA--TAATAGATATATAGTTTG-----TACCTTTTACAGGAAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTACAGGTTTGGCGCAGCGTTTGTTCAGAACG 2135
Qy 1156 GTGACTTTCTGTAACATGCGATGATA--TTAACTTTTATAAAGTT 1203
Db 2136 CCGCCTCCACCCAGATCGCGTATATAGATATGCAATTTATTTTACTT 2186

RESULT 9

US-10-140-923-381
; Sequence 381, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-923-381

Query Match 47.6%; Score 3593; DB 15; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

Qy 1 GATCTGGGCTGTGCTGCAGGAAAAAGCAAAATTC--TGGAAAGTTAATGTTTGA-----GTG 54
Db 191 GCGCCGCGGGGAAACGG-----CGGATTCCTCGCGGTCAACACCACTGATCCCA 244
Qy 55 ATTTTAAATCCTTCTGCG-----GAGAGCCCGCTCTCCCGGTATACG 103
Db 245 AAACATTCATCTC--CCGCGCGCGCGCTGCGAGCGCCCGCAGTCCGCGCCCGCC 303
Qy 104 GCTTCTCATTTTGAATCC--GCG-GTCCGCGGTTCGSCGTC---AGACCAAGCGG 158
Db 304 GC--CTTCCCTGTGCGCTTGCAGCGCTGCG--CACCGCGCCCGAGCCAGCCAG 359
Qy 159 AGGAAGCTTGTGCAATTTAAGCGGGGTGTG-AAAG--CCACAGGCC--GGCGGGCGGG 215
Db 360 AGCGCGCGG--AGCG---GAGCGCGCGAGCTCGTCCGCGCGCGCGCGCGCGCG 413
Qy 216 GCGAGGCGG---GCCATTTGAATAAGA---GGC-----GTGCTTCA 255
Db 414 GCGTAGCGCGCGCGC---TGGATGCGGACCCGCGCGCGGAGAGCGCGCGCGCG 469
Qy 256 GGCAGGCTCTATAAGTGAACCGCGCG-----GCGAGCGTGG 292
Db 470 CGAAGACGACTTTCAGTCCCGGAGCGCGCCCGCCCAACCCCTACGATGAAGAGCGTCCG 529
Qy 293 C-----GC--GTTCAGGTCA-----CTGTAGC---GGACT--TCTTTTGGTT--T 329
Db 530 CTGGAGGAGCGCGCTGCTGTG--CATGGGTGCTGTGCTGCGAGGCTGGCAGCGC 588
Qy 330 TCTTTCTCT---TTGGG---GCACCTCTGACTCA-----CT--CCCCAG 366
Db 589 CAGTCCAGGTGCTGCTGCTGCTACATGAGCCCAAGTCAACAGCTGCCCCAG 648
Qy 367 CA-----TGAGCGCGTGAACCG--GTGCGCGGTGTGACAGGCGGTGTG--CTGCTG- 418
Db 649 CAGGCGCTGCG--GCTGTGCGCGTGGGCATCCCTGTGCTCCAGCAGCGCATCTTCTG 706
Qy 419 TCGGAA-----CGCA---GTCTG--GCCATGCCCGG---GCCAG-----452
Db 707 ACGGCAACCGCATCTCGCATGTGCCAGTGCAGCTGCCAGTTCGTGCTGCGCGCACTCACA 766
Qy 453 -----GGAAG-----GGCCCG-----GCAGCTG-----470
Db 767 TCCTGTGGTGCACCTCGAATGTGCTGGCCCGAATTGATCGCGTGTCTTCACTGGCGCTGG 826
Qy 471 -----AGGAGCGCTG-----AGC-----TTGC-----TGG-----491
Db 827 CCCTCTGGAGCAGCTGGAGCTTCCAGCGATATGACAGCTCCGGTGTGGAGCCCTGCCA 886
Qy 492 -----ACGACATGAACCAAC-----TGCTATCC-----C-GCTTGGCGGAAC 527
Db 887 CATTCACGCGCTGGCGCGCTTACACAGCTGCACCTGGAGCGCTGCGCGCTGCAGGAGC 946

```
QY 528 TGTATCCCGAG-----TCC-----C 543
Db 947 TGG-GCCCGGGGTGTTCCGCGCTCGGCTGCGTACCTCTACCTGCGAGCAAC 1005
QY 544 GAG-----AGCACT-----CAGCTT-----AGCAG-----566
Db 1006 GCCTGCGAGCACTGCTGATGACACCTTCGCGACTGGGCAACTCAGACCTCTTC 1065
QY 567 -----TGAAA-----TCTAC-----AGCGGTGATC-----GACT--ACATCTC 601
Db 1066 CTGACGGCAACCGCATCTCCAGGTGCCGAGCGGCTTCGTTGGTGGCTGACAGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC-----617
Db 1126 GACCGTCTCTACTGCAACCAAGACCGCGTGCGCCATGTGCACCGCATGCTTCGTGAC 1185
QY 618 --TGGCG-----AGCAGC-----630
Db 1186 CTGCGCGCTCATGACACTCTATCTGTTGGCAACAACTATCAGCGTGCCCACTGAG 1245
QY 631 -CCTGACCCCTGATGGCC--CC--ACCT--TC--CCTATCC-----A 666
Db 1246 GCCCTGCGCCCTCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1305
QY 667 GACAGCGAGCTCG-----CTCGGAATGTT-----CATCTCCACGACAAA 709
Db 1306 GACTGCGGGACGCCCACTCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1365
QY 710 AGGAGCTTTTCC-----ACTGACTGCGCGTGTCT--GACACT-----CCA- 751
Db 1366 GTGCGCTGACGCTCCGCAACGCTGCT--GGCGTGACCTCAACGCTAGCTGCCAA 1424
QY 752 -GAAC-CAAG--TGC-----TGCG-----GCCGT-----TCTG--C--CTGGGA 784
Db 1425 TGACCTGAGGCTGCGCTGTGTCACCGGCTTACCATCCATCTGACCGGAGGC 1484
QY 785 CCGCG-----GGAACCTCT-----CCTCGCGAAGCG--GAC--GG- 817
Db 1485 CACCGATGAGAGCGCTGCGGCTTCCAACTGTGCGCAGCAAGTCCGCTGACAAGGC 1544
QY 818 --GAGGATGG--GCC-----CCAATCTG-----838
Db 1545 CTCAGTACTGAGCTGGAACACGAGCTTGGGAGGCAATGCGTGAAGGAGCGGTGCC 1604
QY 839 --CC-----CTGCGCACTTGACTTC-----ACCAATCCCTTC-----T 871
Db 1605 GCCCGTGACAGCGCGCGGCAACGGCTCTGGCCCAAGGCAATCAATGACTCACCTT 1664
QY 872 GGAGCTAACTGG-----TGCTCA-----GGAGCGAAG-AGTGT 907
Db 1665 TGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1724
QY 908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGCTAGCTC 939
Db 1725 GCCACAGGTTCCCACTCGGCGCTCGCGGA--GGCAGGCTGTTCAAGCAAGAAC 1782
QY 940 TG-----GCCA--CCAGCTGGG-----GAGC-----959
Db 1783 CACACCGCAGCACTGCTGCTGCGGCAAGCAGGAGCGGGGTGCGGAGTGTGTGAC 1842
QY 960 -----TCA-----CCTGTCT-----CCACCCCACTCC-----982
Db 1843 TCAGAGGCTCAGTGGCTTACCCAGCTCAGCTGAGCTCAACCCCTGCGGCTGGCG 1902
QY 983 CAAGTCTTAAGTCTTTTC-----AGAGCTG-----1009
Db 1903 CTGGTGTCTGACAGTCTTGGGCTGCTGACCCCGGAGCGGACACAGAGCTGTCTCA 1962
QY 1010 -----GAGTGTG-----GAAGAGTGTGCTCTCCAACTATGCCAAGGC-----GG 1053
Db 1963 GCAGCCAGTGTGTATATACAGGGT---CT-CTCTCAAGC---CGCCAGCAGCGCGG 2017
```

```
QY 1054 CGGAGA-----GCTGGTCT--TCTGGT-CTCCTTGAGAAAGTCTCTGTGCGCTG 1102
Db 2018 CGGCGGACCGGTGGGCGAGCCAGCCAGGTCTCTCCCTGATGGAG--CCTGCGCGCGC 2075
QY 1103 ATTTATGACTCTA--TAATAGAGTATATAGTTTTC-----TACCTTTTTCAGGAAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTTCGCGCAGCGTTTGTTCAGAACG 2135
QY 1156 GTGACTTCTTGTATACAAATGCGATATA--TAACTTTTATAAAGTT 1203
Db 2136 CGGCTCCACCCAGATCGCGTATATAGATATGCAATTTTATTACTT 2186
```

RESULT 10

US-10-141-756-381

; Sequence 381, Application US/10141756

; Publication No. US20030207359A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DePorge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330RIC200

; CURRENT APPLICATION NUMBER: US/10/141,756

; CURRENT FILING DATE: 2002-05-08

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 381

; LENGTH: 2236

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-141-756-381

Query Match

47.6%; Score 3593; DB 15; Length 2236;

Best Local Similarity 39.8%; Pred. No. 6.2e-171;

Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

```
QY 1 GATCTGGGGTGTGCGCAGAAAGCAATTC--TGGAGTTATGTTTGA----GTG 54
Db 191 GCGCGCGGGGGAACCGG-----GCGGATTCCTGCGCGTCAACACCACTGATCCCA 244
QY 55 ATTTTAAATCCTTCTGCGG-----GAGAGCGCGGCTCTCCCGGTATCAGC 103
Db 245 AAACATTATCTCT--CCGCGCGCGCGCGCTGCGAGCGCGCGGAGTCCGCGCGCGCGC 303
QY 104 GTTCTCTCATTTTGAATCC--GCG-GTCCGCGGTCTTCGGGTC---AGACAGCGCGG 158
Db 304 GC--CTCGCCCTGTGCGCCCTGCGCGCCCTGCG--CACCGCGCGCGCGAGCCAGCCAG 359
QY 159 AGGAGACCTGTTTGAATTTAAGCGGGTGTG-AAACG-CCACAGGGCC--GGCGGGGGCGG 215
Db 360 ACCCGGGCGG--AGCG-----GAGCGCGCGAGCTCTGTCGCGCGCGCGCGCGCGG 413
QY 216 GCCGAGGGCGG-----GCCATTTTGAATAAGA---GGC-----GTGCTTCCA 255
Db 414 GCGGTAGCGCGCGCGC-----TGGATGCGGACCCCGCGCGCGGAGAGCGGCGCGCGCC 469
QY 256 GCGAGGCTCTAAGTAGCGCGCGG-----GCGAGCGTGGC 292
```

Db 470 CGAAACGACTTTCAGTCCCGCGCGCGCCGCGCCACCCCTACGATGAAGAGCGGCTCG 529
Qy 293 C-----GC--GTTCAGGTCA-----CTGTAGC-----GGACT--TCTTTTGGTT--T 329
Db 530 CTGAGAGGAGCGCGCTGCTGG--CATGGGTGCTGTGGCTGCGAGCCTTGGCAGGTGCGAGCC 588
Qy 330 TCTTTCTCT-----TTGGG--GCACCTCTGGAATCA-----CT--CCCCAG 366
Db 589 CCAATGCCAGGTGCTGCTGATGCTACATGAGCCCAAGGTGACGACAGAGTGCCTCCCGCAG 648
Qy 367 CA-----TGAAGGCGGTGAGCCCG--GTGCGCGGCTGTACGAGCGGTGTG--CTGCTG-- 418
Db 649 CAGGCGCTGCG--GCTGTGCGGCTGCGCATCCTGCTGCGACGCGCATCTTCTCTG 706
Qy 419 TCGGAA-----CCCA--GTCTG--GCCATGCGCCCGG--GCCGAG----- 452
Db 707 ACGGCAACCGCATCTCGAATGTGCCAGTGCAGCTTCCGTGCTGCGGCAACCTCACCA 766
Qy 453 -----GGAAG-----GCCCG-----GAGCTG----- 470
Db 767 TCTGTGGCTGCACTCGAATGTGTGCGCCGAAATGATGCGGCTGCTTCACTGGGCTG 826
Qy 471 -----AGGAGCGCTG-----AGC-----TTGC-----TGG----- 491
Db 827 CCTCTGAGAGCACTGCACTAGCGATANTGACAGCTCCGCTGTGTGGACCTTGCCA 886
Qy 492 -----ACGACATGAACCA-----TGCTACTCC-----C-GCCTGCGGGAAC 527
Db 887 CATTCACGGCTGCGCGCGCTACACAGCTGCACCTGGACGCTGCGGCTGCGGCTGCGAGG 946
Qy 528 TGGTACCGGAG-----TCC-----C----- 543
Db 947 TGG-GCCCGGGGCTGTTCCCGGCGCTGCGGCTGCGCTGCGAGTACCTTACCTGCGAGCAAC 1005
Qy 544 GAG-----AGGCAC-----CAGCTT-----AGCCAGG----- 566
Db 1006 GCGGTGAGGACTGCTGTGATGACACCTTCCGCGACCTGGGCAACCTCACACACCTCTTC 1065
Qy 567 -----TGAAA-----TCTTAC-----AGCGGTATC-----GACT--ACATCTC 601
Db 1066 CTGACGCGCAACCGCATCTCCAGGCTGCGGAGCGCGCTTCCGTGGGCTGCAACAGCCTC 1125
Qy 602 GAC-----CTGCA-----GGTAGTCC----- 617
Db 1126 GACGCTCTCTACTGACCAAGAACCGGCTGCGCCATGTGACCGCATGCTTCCGTGAC 1185
Qy 618 -----TGCCCG-----AGCCAGC----- 630
Db 1186 CTGGCGCCTCATGACACTCTATCTGTTGCCAACAATCTATCAGCGCTGCGCCACTGAG 1245
Qy 631 -CCCTGAGACCCCTGATGCGCC--CC--ACCT--TC-----CCATCC-----A 666
Db 1246 GCGCTGGCGCCCTGCTGCGCTGAGTACCTGAGGCTCAAGCAACCCCTGGGTGTGT 1305
Qy 667 GACAGCCGAGTCTG-----CTCGGAACCTGT-----CATCTCAACGACAAA 709
Db 1306 GACTCGCGGCAAGCCCACTGCGGCTGCGTGCAGAAATTCGCGGCTCTCTCTCCGAG 1365
Qy 710 AGGAGCTTTTGC-----ACTGACTGCGGCTGCTCT--GACACCT-----CCA- 751
Db 1366 GTGCGCTGAGGCTTCCGCAACGCTGGCT--GGCGTGACCTCAAGCCCTAGTGCAAA 1424
Qy 752 -GAAC-GCAGG--TGC-----TGGC-----GCCGCT-----TCTG--C-CTGGGA 784
Db 1425 TGACCTGAGGCTGGCTGTGGCCACCGGCCCTTACCATCCCATCTGGACCGGCGAGGC 1484
Qy 785 CCGCG-----GGAACCTCT-----CTGCGGAAGCG--GAC--GG- 817
Db 1485 CACCGATGAGGAGCGCGCTGGGCTTCCCAAGTGTGCCAGCGCATGCGCTGCAAGGC 1544
Qy 818 --CAGGATGG--GCC-----CCAACCTCG----- 838

Db 1545 CTCAGTACTGAGGCTTGAAGACACAGCTTCGGCAGGCAATGCGCTGAAGGACCGGTGCC 1604
Qy 839 --CC-----CTGCCCACTTGACTTC-----ACCAAATCCCTTCC-----T 871
Db 1605 GCGCGGTGACAGCCCGCGGCAACGGCTCTGGCCCAAGGCAATCAATGACTCACCTT 1664
Qy 872 GGAGACTAAACTGG-----TGCTCA-----GGAGCGAGG--ACTGT 907
Db 1665 TGGGACTCTGCTGCTGCTGTGTGAGCCCGCTCCTCAGTGCAGTGGCGCGGAGGCTCGA 1724
Qy 908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGTAGTCTC 939
Db 1725 GCCACCAAGGTTCCTCCACTCGGCGCTCGCGGA--GGCCAGGCTGTTCACGCAAGAAC 1782
Qy 940 TG-----GCCA--CCAGCTGGG-----GAGC----- 959
Db 1783 GCGACCCGACGCACTGCCGTCTGGGCCAGGACGCGGGGTGGCGGACTGGTGAC 1842
Qy 960 -----TCA-----CCTGCTC-----CCACCCACCC----- 982
Db 1843 TCAGAGGCTCAGTGGCCCTACCCAGCTCACCCTGACGCTCACCCCTGGGCTGGCG 1902
Qy 983 CAAGTCTTAAGTCTTTTC-----AGAGCGTG----- 1009
Db 1903 CTGGTGTGTGACAGTGTCTGGGCCCTGTGACCCCGAGCGGACACAAGAGCGTGTCA 1962
Qy 1010 -----GAGGTG-----GAAGGAGTGGCTGCTTCCAACTATGCCAAGC-----GG 1053
Db 1963 GCGACCGAGTGTGTATCATAGGGT--CT-CTCTCCAGC--CGCCAGCCAGCCGG 2017
Qy 1054 CGCGAGA-----GCTGGTCT--TCTGGT--CTCTTGGAGAAAGTCTGTTCCTCTG 1102
Db 2018 CGGCGACCCGTGGGCGAGGCGCAGGCTCTCTCTGATGGAGC--CTTGGCGCGCGC 2075
Qy 1103 ATTTATGACTCTA--TAATAGATATATAGTTTG-----TACCTTTTATAGAGGAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTTGGCGGCGAGGTTTGTTCAGAACG 2135
Qy 1156 GTGACTTCTTGAACAATGCGATGTATA--TTAACTTTTATATAAAAGTT 1203
Db 2136 CGGCTTCCACCCAGATCGGCTATATAGATATGCAATTTTACTT 2186

RESULT 11

US-10-141-759-381
; Sequence 381, Application US/10141759
; Publication No. US20030207361A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Guirney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P33303R1C197
; CURRENT APPLICATION NUMBER: US/10/141,759
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381

; LENGTH: 2236		
; TYPE: DNA		
; ORGANISM: Homo Sapien		
; US-10-141-759-381		
Query Match 47.6%; Score 3593; DB 15; Length 2236;		
Best Local Similarity 39.8%; Pred. No. 6.2e-171;		
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;		
QY	1	GATCTGGGTCCTGCGAGGAAAGCAATTC--TGGAGTTAATGGTTTGA-----GTG 54
DB	191	GGCGCGGGGGAACCGG-----GCGATTCTCGCGGTCAAAACCACTGATCCCAT 244
QY	55	ATTTTAAATCCCTTGCTGGCG-----GAGAGGCCCGCTCTCCCGGGTATCAGC 103
DB	245	AAACATTCATCTTC--CGGCGCGCGCGGTGCGAGCGCCCGCAGTCCGCGCGCGCC 303
QY	104	GCTTCCTCATCTTTGAATCC--CGG--GCTCCGGGTCCTTCGGGTC---AGACCGCGG 158
DB	304	GC--CCTCGCCCTGTGCGCCTCGCGCCCTGCG--CACCCGCGCGCGAGCCGACCCAG 359
QY	159	AGGAGCGCTTTTGAATTTAAGCGGCTGTG--AAG--CCAGGGGC--GCGCGGGCGCGG 215
DB	360	AGCGGGCGG--AGCG---GAGGCGCGAGCTGTCTCCGCGCGCGGCGCGCGCGG 413
QY	216	GCCGAGCGCG---GCCATTTTGAATAAGA---GGC-----GTGCTTCCA 255
DB	414	GCCGTAGCGCGCGCGC---TGGATGCGGACCCCGCGCGGAGACGCGCGCGCGCC 469
QY	256	GGCAGGCTCTAATAGTACCGCGCG---GCGAGCGCTGGCG 292
DB	470	CGAAACACATTTTCACTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
QY	293	C-----GC--GTTCAGGTCA-----CTGTAGC---GACT--TCTTTGGTT--T 329
DB	530	CTGAGGAGCGCGGCTGCTG--CATGGTGTCTGTGCTGCGAGGCTTGGCAGGTGGCAGCC 588
QY	330	TCCTTCTCT-----TTGGG--GCACCTCTGGACTCA-----CT--CCCCAG 366
DB	589	CCATGCCAGGTGCTCGTATGCTAATGAGCCCAAGTGAAGAGAGGCTGCCCGCAG 648
QY	367	CA-----TGAAGGCGCTGAGCGCG--GTGCGCGGCTGTACGAGCGGTTGT--CTGCTG-- 418
DB	649	CAGGCGCTGCAG--GCTGTGCGCGTGGGCATCCCTGTGCGCAGCGCATCTTCTCTG 706
QY	419	TCGGAA-----CGCA---GTCTG--GCCATCGCGCGG--CCCGAG----- 452
DB	707	ACGGCAACCGCATCTCGCATGTGCCAGTGCAGCTTCCGTCGCTGCGCGCACTCACCA 766
QY	453	-----GGAAG-----GGCCCG-----GCAGCTG----- 470
DB	767	TCCTGTGCTCACTCGAATGTGTGCGCGCGAATTGATGCGGTTCCTTCACTGGCGCTG 826
QY	471	-----AGGAGCGCTG-----AGC-----TTGC-----TGG----- 491
DB	827	CCCTCCTGGAGCAGTGGACCTCAGCATATGCAAGCTCCCGTCTGTGGAACCTTGCCA 886
QY	492	-----ACGACATGAACCA-----TGCTACTCC-----C-GCCTCGGGGAAC 527
DB	887	CATTCACGCGCTGGCGCGCTACACAGCTGACCTGGACCGCTGCGGCTGCAGGAGC 946
QY	528	TGGTACCCGAG-----TCC-----C 543
DB	947	TGG--GCCCGGGCTGTTCCCGCGGCTGGCTGCCCTGCACTTCTTACCTGCAGGAAAC 1005
QY	544	GAG-----AGGCACCT-----CAGCTT-----AGCCAGG----- 566
DB	1006	GCGCTGCGAGCACTCCCTGTATGACACCTTCGCGACCTTGGCAACCTCACACACCTTTC 1065
QY	567	-----TGGAAA-----TCCTAC-----AGCGGTCTATC-----GACT--ACATCTC 601
DB	1066	CTGCAGCGCAACCGCATCTCCAGGTGCGCGCGCGCTTCCGTGGGCTGCACAGCCTC 1125

QY	602	GAC-----CTGCA-----GGTAGTCC----- 617
DB	1126	GACGCTCTCTCTACTGTGACACAGAACCGGTGGCCCATGTGTGACCCCGCATGCTTCCGTGAC 1185
QY	618	--TGGCG-----AGCCAGC----- 630
DB	1186	CTTGGCGCGCTCATGACACTCTATCTGTGTTTGGCAACAATCTATCAGGCTGCCACTGAG 1245
QY	631	--CCCTGACACCCCTGTATGGGCC--CC--ACCT-----TC-----CCATCC-----A 666
DB	1246	GCCCTGGCCCCCTGCGTGCCTGAGTACCTGAGGCTCAACGACAAACCCCTGGTGTGT 1305
QY	667	GACAGCCGAGTGG-----CTCCGGAAGTTGT-----CATCTCCACGACAAA 709
DB	1306	GACTGCGCGGACGCCCACTCTGGGCTGGCTGAGAGTTCCGGGCTCTCTCTCGAG 1365
QY	710	AGGAGCTTTTGGC-----ACTGACTGCGCGGTGTCT--GACACCT-----CCA-- 751
DB	1366	GTGCGCTGCAAGCTCCCGCAACGCTGGCT--GGCGGTGACCTCAAAAGCGCTAGTGGCAA 1424
QY	752	--GAAC--GAGG--TGC-----TGGC-----GCCGT-----TCTG--C--CTGGGA 784
DB	1425	TGACTGCAAGGCTGCGCTGTGGCCACCGGCTTACCATCCCATCTGGACCGGCGAGGC 1484
QY	785	CCCCG---GGAACCTCT-----CCTGCCGGAAGCG--GAC--GG-- 817
DB	1485	CACGATGAGGAGCGCTGGGCTTCCCAAGTGTGCGCAGCGAGATGCGCTGACAAGGC 1544
QY	818	--CAGGGATGG--GCC-----CCAACTTCG----- 838
DB	1545	CTCAGTACTGAGGCTTGGAAAGCCAGCTTGGCAGGCAATGCGTGAAGGACGCGTGCC 1604
QY	839	--CC-----CTGCCCACTTGACTTC-----ACCAATCCCTTTC-----T 871
DB	1605	GCCGCTGACAGCCCGCGGCAACGGCTCTGGCCCGCAGGCACATCAATGACTCACCTT 1664
QY	872	GGAGACTAAACCTCG-----TGCTCA-----GGAGGGAAGG--ACTGT 907
DB	1665	TGGACTCTGCTGGCTGTGCTGAGCCCGCTCACTGCAATGCGCGCGCGGCTCCGA 1724
QY	908	G-----AACTTG-----TAGCTGAAGAGCCAG-----AGTAGCTC 939
DB	1725	GCCACCGGTTCCCGACCTCGGCGCTGCGCGA--GGCAGGCTGTTCAAGCAAGAAC 1782
QY	940	TG-----GCCA--CCAGCTGGG-----GAGC----- 959
DB	1783	CGCAGCGCAGCCACTGCGCTTGGCGCAGGCGAGCGGGGTGGCGGACTGTGTGAC 1842
QY	960	-----TCA-----CCCTGCTC-----CCACCCCACTCC----- 982
DB	1843	TCAGAGGCTCAGTGGCGCTTACCAGCCTACCTGAGCCTCACCCCTGGGCTGGCG 1902
QY	983	CAAGTTCTAAGGCTCTTTTC-----AGAGCGTG----- 1009
DB	1903	CTGTGTCTGTGGACAGTGTGTTGGCGCTGTGACCCCGCAGCGGACACAAGAGCGTGTCA 1962
QY	1010	-----GAGGTGTG-----GAAGGATGGTGTCTTCCAACTATGCCAAGGC-----GG 1053
DB	1963	CGAGCCAGGTGTGTATACATACGGGT--CT--CTCTCCACGC--CGCCAGCCAGCGGG 2017
QY	1054	CGGCAAG-----GCTGGTCT--TCTGGT--CTCCTTGGAGAAAGTTCTGTGCGCTG 1102
DB	2018	CGGCGGACCCGTGGGCGAGCGCAGCCAGGTCTCTCCCTGATGAGC--CCTGCGCGCGC 2075
QY	1103	ATTTATGACCTTA--TAATAGATATATAGTTTG-----TACCTTTTATCAGGAAG 1155
DB	2076	CACCCCATCTCCACCCCATCATGTTTACAGGTTTCGGCGGCGAGGTTTGTTCAGAACG 2135
QY	1156	GTGACTTCTGTAAACAATGCGATGTATA--TTAACTTTTATAAAAGTT 1203
DB	2136	CGGCTTCCCAACCGATCGCGGTATAGAGATATGCAATTTTATTTACTT 2186

RESULT 12	
US-10-140-805-381	
; Sequence 381, Application US/10140805	
; Publication No. US20030207417A1	
; GENERAL INFORMATION:	
; APPLICANT: Baker, Kevin P.	
; APPLICANT: Beresini, Maureen	
; APPLICANT: DeForge, Laura	
; APPLICANT: Desnoyers, Luc	
; APPLICANT: Filvaroff, Ellen	
; APPLICANT: Gao, Wei-Qiang	
; APPLICANT: Gerritsen, Mary E.	
; APPLICANT: Goddard, Audrey	
; APPLICANT: Godowski, Paul J.	
; APPLICANT: Gurney, Austin L.	
; APPLICANT: Sherwood, Steven	
; APPLICANT: Smith, Victoria	
; APPLICANT: Stewart, Timothy A.	
; APPLICANT: Tamas, Daniel	
; APPLICANT: Watanabe, Collin K	
; APPLICANT: Wood, William	
; APPLICANT: Zhang, Zemin	
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC	
; FILE OF INVENTION: ACIDS ENCODING THE SAME	
; FILE REFERENCE: P3330R1C176	
; CURRENT APPLICATION NUMBER: US/10/140.805	
; CURRENT FILING DATE: 2002-05-07	
; Prior Application removed - See File Wrapper or Palm	
; NUMBER OF SEQ ID NOS: 550	
; SEQ ID NO 381	
; LENGTH: 2236	
; TYPE: DNA	
; ORGANISM: Homo Sapien	
US-10-140-805-381	
Query Match 47.6%; Score 3593; DB 15; Length 2236;	
Best Local Similarity 39.8%; Pred. No. 6.2e-171;	
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;	
Qy	1 GATCTGGGCTGTGCGAGGAAAGCAAAATTC--TGGAAGTTAATGTTTGA----GTG 54
Db	191 GCGCGCGGGGAACCGG-----GCGGATTCCTCGCGCTCAAAACACCTGATCCCAT 244
Qy	55 ATTTTAAATCTTCTCTGGG-----GAGAGCGCGCTCTCCCGGATCAGC 103
Db	245 AAACATTCATCTC--CGGCGCGCGCGCTCGAGCGCCCGCAGTCCGCGCGCGCC 303
Qy	104 GTTCTCTATTCTTTGAATCC--GCG--GCTCCGCGCTCTTCGGCGTC---AGACGAGCGG 158
Db	304 GC--CTTCGCTGTGGCCCTGCGCGCTGCG--CACC CGCGCGCGAGCCAGCCAG 359
Qy	159 AGGAAGCCTGTTTGAATTAAGCGGCTGTG-AAAG--CCAGGGCC--GGCGGGGGGG 215
Db	360 AGCGGGCGG--AGCG--GAGCGCGCGAGCTCTGTCGCGCGCGCGCGCGCGCGCG 413
Qy	216 GCGAGGCGG--GCCATTTTGAATAAGA--GCG-----GTGCTTCCA 255
Db	414 GCGTAGCGCGCGCGCC-----TGGATGCGGACCCCGCGCGCGGAGCGCGCGCGCC 469
Qy	256 GCGAGGCTCTATAAGTGACCGCGCG-----GCGAGGCTGCG 292
Db	470 CGAAACGACTTTCAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
Qy	293 C-----GC--GTTCAGGTCA-----CTGTAGC-----GGACT-TCTTTTGGTT--T 329
Db	530 CTGAGGAGCGCGCTGTGTG--CATGGGTGCTGTGTGTGCTGAGGCTTGGCAGGTGCAGCC 588
Qy	330 TCTTTCTCT-----TTGGG---GCACCTCTGACTCA-----CT--CCCCAG 366
Db	589 CCATGCCAGGTGCTCGGTATGCTACATGAGCCCAAGGTACGACAAAGTGTGCCCGCAG 648
Qy	367 CA-----TGAGCGCTGAGCCCG--GTGCGCGGCTCTACGAGGCGGTGTG--CTGCTG- 418

Db	649 CAGGCGCTGCGAG--GCTGTGCCCGTGGGCAATCCCTGTGTCGCGAGCGGCAATCTTCCTGC 706
Qy	419 TCGGAA-----CGCA--GTCGTG--GCATCGCCCGG--GCGGAG-----452
Db	707 ACGCAACGGCAATCTCGCATGTGCAGGTGCGAGTTCGCTGCGCGCAACCTCACCA 766
Qy	453 -----GGAAG-----GGCCCG-----GCAGCTG-----470
Db	767 TCCTGTGGCTGCACTCGAATGTGCTGGCCCGAATGATGCGGCTGCTTCACTGCGCTGG 826
Qy	471 -----AGAGCGGCTG-----AGC-----TTGC-----TGG-----491
Db	827 CCCTCTGAGCAGCTGGACCTCAGCATATGACAGCTCCGGTCTGTGAGCCCTGCGCA 886
Qy	492 -----ACGACATGAACCAC-----TGCTACTCC-----C-GCCTGCGGGAAC 527
Db	887 CATTCCACGGCTGGCGCGCTTACACACGCTGACACGCTGACCGCTGCGGCTGCGAGGAGC 946
Qy	528 TGGTACCCGGAG-----TCC-----C-----C 543
Db	947 TGG-GCCCCGGGCTGTTCGCGCGCTGCTGCGCTGCGCTGCTTCACTGCGAGGACAAAC 1005
Qy	544 GAG-----AGGCACT-----CAGCTT-----AGCCAGG-----566
Db	1006 GCGCTGCGGCACTGCTGATGACACCTTCGCGACCTGGGCAACTCAGACACCTCTTTC 1065
Qy	567 -----TGGAAG-----TCCTAC-----AGCGGTATC-----GACT--ACATTCTC 601
Db	1066 CTGACGCGCAACCGCATCTCCAGCGTGGCCGAGCGCGCTTCCGTGGGTGACACGCTTC 1125
Qy	602 GAC-----CTGCA-----GGTAGTCC-----617
Db	1126 GACGCTCTCTACTGACAGAACCGGTGGGCCCATGTGACCCGCGATGCTTCCGTGAC 1185
Qy	618 --TGCCCG-----AGCCAGC-----630
Db	1186 CTGGCGGCTCATGACACTCTATCTGTTGGCAACAATCTATGAGCGTGCCCACTGAG 1245
Qy	631 -CCCTGAGCCCTGATGCCCC--CC--ACCT--TC-----CATCC-----A 666
Db	1246 GCCTTGGCCCCCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
Qy	667 GACAGCGAGCTG-----CTCCGGAACCTGT-----CATCTCCAAACGAAAA 709
Db	1306 GACTGCGGCGCAGCCCACTCTGGGCTGGCTGCGAGAGTTCGCGGCTCTCTCTCCGAG 1365
Qy	710 AGAGCTTTTGGC-----ACTGACTCGCGCTGCT--GACACT-----CCA- 751
Db	1366 GTCCCTTGAGCGCTCCCGCAACGCGCTGGCT--GGCCGTGACCTCAAAAGCTAGTCCCA 1424
Qy	752 -GAAC-GCAGG--TGC-----TGGC-----GCCCGT-----TCTG--C--CTGGGA 784
Db	1425 TGACCTGCGAGGCTGGCTGTGGCCACCGGCTTACCCTCCATCTGGACCGGAGGCG 1484
Qy	785 CCCCC-----GGAACCTCT-----CTTCCGCGGAAGCCG--GAC--GG- 817
Db	1485 CACCGATGAGGAGCGCTGGGCTTCCCAAGTGTCTCCAGCCAGATGCGCTGACAAGGC 1544
Qy	818 --CAGGATGCG-GCC-----CCAACTTCG-----838
Db	1545 CTGACTGAGGCTGAGAGACAGCTTCGGCAGGCAATCGCTGAAGGAGCGCGTGC 1604
Qy	839 --CC-----CTGCCCACTTGACTTC-----ACCAATCCCTTC-----T 871
Db	1605 GCGCGGTGACAGCCCGCGGCAACGCTTGTGCGCCCAACGCAATCAATGACTCAACCTT 1664
Qy	872 GGAGACTAATACCTGG-----TGCTCA-----GGAGCGAAGG-ACTGT 907
Db	1665 TGGGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1724
Qy	908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGCTAGCTC 939
Db	1725 GCCACGAGGCTTCCCACTCGGCGCTCGCCGGA--GGCCAGGCTGTTTCAACGCAAGAC 1782

940	TG-----GCCA--CCAGCTGGGC-----GACG-----	959
1783	CGCACCCGACGACCATCGCCGCTTGGCCAGGCAGGCAGCGGGGTGGCGGACTGGTGAC	1842
960	-----TCA--CCCTGTCT-----CCACCCCAACCC-----	982
1843	TCAGAAGGCTCAGGTGCCCTACCCAGCCTCACTGCAGCCTCACCCCTCGGCCCTGGCG	1902
983	CAAGTTCNAAGTCTTTTC-----AGACGTG-----	1009
1903	CTGTGCTGTGGACAGTGTCTGGCCCTGCTGACCCCGACGACACAGAGCGTGCTCA	1962
1010	-----GAGGTG-----GAGGAGTGCTGCTCCAAACTATGCCAAGC-----GG	1053
1963	GCAGCCAGGTGTGTACATACGGGT--CT-CTCTCCACGC--CGCCAAAGCAGCCGGG	2017
1054	CGGCAGA-----GCTGGTCT--TCTGGT-CTCCTTGGAGAAAGTTCTGTGGCCCTG	1102
2018	CGGCCGACCGTGGGCAGGCAGCCAGGTCTCCTGATGGACG--CCTCGCGCCCG	2075
1103	ATTATGAACCTCA--TAATAGAGTATATAGTTTTG-----TACCTTTTTCAGGAG	1155
2076	CACCCCACTCCACCCCATCATGTTTACAGGGTTCGGCGCAGCGCTTGTTCAGAAAG	2135
1156	GTGACTTTCGTAAACAATGGATGTATA--TTAACTTTTATAAAGTT	1203
2136	CGGCTTCCCAACCCAGATCGGGTATATAGAGATATGCATTTATTTTACTT	2186

```

RESULT 13
US-10-140-864-381
; Sequence 381, Application US/10140864
; Publication No. US20030207419A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Tumanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C184
; FILE REFERENCE: P330R1C184
; CURRENT APPLICATION NUMBER: US/10/140,864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-864-381

```

```

Query Match      47.68; Score 3593; DB 15; Length 2236;
Best Local Similarity 39.88; Prod No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY      1  GATCTGGGGTCTGCCAGGAAAAAGCAATTC-----TGGAACTTATGTTTGA-----GTG  54
Db      191  GCGCGCGGGGGAACCGG-----GCGGATCTCTCGGGCGTCAACACCACTGATCCCAT  244

```

QY	55	ATTTTAAATCTTTGCTGGCG-----GAGAGCGCCGCTCTCTCCCGGTATCAGC	103
Db	245	AAACATTATCTCTC-CCGGCGCCCGCTGCGAGCGCCCGCGCAGTCCGCGCGCGCC	303
QY	104	GCCTTCCTCATTTCTTTGAATCC-GCG-GTCCGCGGTCTTTCCGGCTC---AGACACAGCGG	158
Db	304	GC---CTTCGCCCTGTGCGCCCTGCGGCCCTTGGC---CACCCGCGCCGAGCCACAGCAG	359
QY	159	AGGAAGCCTCTTTTGCAATTTAAGCGGGGTGTG-AAAG-CCAGAGGCC-GGCGGGGGGGGG	215
Db	360	AGCGCGCGG---AGCG---GAGCGCGCGAGCCTCGTCCGCGCGCGCGCGCGCGCGG	413
QY	216	GCCGAGGCGG---GCCATTTTGAATAAGA---GGC-----GTGCTTCCCA	255
Db	414	GCCGTAGCGCGCGCGCC---TGGATGGAGACCCGCGCGCGGGAGAGCGGCGCGCGCC	469
QY	256	GGCAGGCTCTATAAGTGAACCGCGCG---GCGAGCGTGTGCG---GCGAGCGTGTG	292
Db	470	CGAAACGACTTTCAGTCCCGGACGCGCGCGGCCCAACCCCTACGATGAAGAGGCGTCCG	529
QY	293	C-----GC-GTTGCAGGTCA-----CTGTAGC---GGACT-TCITTTTCGTT--T	329
Db	530	CTGAGGAGACCGCGCTGTGTG-CATGGGTGTGTGGCTGCGAGGCTGTGCAAGTGGCAGCC	588
QY	330	TCITTTCTCT-----TTGGG---GCACCTCTGCACTCA-----CT---CCCCAG	366
Db	589	CCATGCCAGGTGCTCTCGTATGTACTAATGAGCCCAAGTGAAGCAAGCTGTGCCCCAG	648
QY	367	CA-----TGAAGCGCTGAGCCCG-GTGGCGGCTGTCTACGAGGCGGTGTG-CTGGCTG-	418
Db	649	CAGGGCTCGAG--GCTGTGCCCGTGGGATCCCTGTGCGCAGCAGCGCATCTCTCTGC	706
QY	419	TCGGAA-----CGCA---GTCTG-GCCATGCCCGCGG---GCCGAG-----	452
Db	707	ACGGCAACCGCATCTCGCATGTGCGCAGTSCCAGCTTCGTGCGTCCGCAACCTCACCA	766
QY	453	-----GGAAG-----GGCCCG-----GCAGCTG-----	470
Db	767	TCCTGTGGCTGCACTCGAATGTGCTGGCCCGCAATTGATGCGGCTGCTTCACTGCGCTGG	826
QY	471	-----AGGAGCGCTG-----AGC---TTGG-----TGG-----	491
Db	827	CCCTCTGGAGCAGCTGAGACTCAGGATATGACAGCTCCGGTCTGTGACCCCTGCCA	886
QY	492	-----ACGACATGAACCA-----TGCTACTCC-----C-GCTTCGGGGAAC	527
Db	887	CATTTCACGGCTGGGCGCGCTACACACGCTGCACTGGAACCGCTGCGGCTGTGAGGAGC	946
QY	528	TGTTACCCGGAG---TCC-----TGCTACTCC-----C 543	543
Db	947	TGG-GCCCGGGGTGTTCCGCGGCTGGCTGCGCTGTGAGTACCTCTACTGCGAGCACAC	1005
QY	544	GAG---AGGCACT-----CAGCTT---AGCCAGG-----	566
Db	1006	GCGGTGCAGCACTGCTGATGACACTTCCGCGACCTGGGCAACCTCACACACCTCTTC	1065
QY	567	-----TGGAAA-----TCCTAC-----AGCGGTATC---GACT-ACATTTCTC	601
Db	1066	CTGACGGCAACCGCATCTCCAGCGTGGCGAGCGCGCCTTCCGTGGGTGACAGCCTC	1125
QY	602	GAC-----CTGCA-----GGTAGTCC-----	617
Db	1126	GACGCTCTCTACTGCAACAGACCGGTGGGCCCATGTGCAACCGCATGCTTCCGTGAC	1185
QY	618	--TGGCGC-----AGCCAGC-----	630
Db	1186	CTTGGCGCCTCATGCACTCTATCTGTTTGGCAACAATCTATACGCGTGCCCACTGAG	1245
QY	631	-CCCTGACCCCTGATGGCC-CC-ACCT-----TC---CCATCC-----A	666
Db	1246	GCCCTGCCCCCTGTGCTGCTCTGAGTACCTGAGCTCAACGACAAACCCCTGGGTGT	1305
QY	667	GACAGCGAGCTCG-----CTCCGGAACTGT-----CATCTCGAACACAAA	709

```
Db 1306 GACTGCGGGCAGCCCACTCTGGGCTGGTGCAGAAAGTTCCGGGCTCTCTCCGAG 1365
Qy 710 AGGAGCTTTTGGC-----ACTGACTGGCGGTGCT--GACACCT-----CCA- 751
Db 1366 GTGCCCTGCAGCTCCCGCAAGCGCTGGCT--GGCGTGAAGCTCAAGCGCTAGCTGCCAA 1424
Qy 752 -GAAC-GCAGG--TGC-----TGGC-----GCCGT-----TCG-----C--CTGGGA 784
Db 1425 TGACCTGCAGGGCTGGCTGGGCGACCGGCTTACCATCCATCTGGACCGCGCAGGC 1484
Qy 785 CCGCG--GGAACCTCT-----TGGC-----GCCGT-----TCG-----C--CTGGGA 817
Db 1485 CACGATGAGGAGCGCTGGGCTTCCAGTGTGCGACGAGATGCCCTGACAGGC 1544
Qy 818 --CAGGATGG-GGC-----CCAAGTCTG-----TGGC-----GCCGT-----TCG-----C--CTGGGA 838
Db 1545 CTCAGTACTGAGCGCTGGAACAGCTTGGGCGACCGCTTGGGCGACCGCTGGAAGGAGCGCGTGC 1604
Qy 839 --CC-----CTGCCCACTTGACTTC-----ACCAATCCCTTCC-----T 871
Db 1605 GCCCGGTGACAGCGCGCGGCAACGCGCTGGGCGACCGCTGGAAGGAGCGCGTGC 1664
Qy 872 GGAGACTAAACCTGG-----TGCTCA-----GGAGCGAAGG-AGTGT 907
Db 1665 TGGGACTGTGCTGGCTGTGCTGAGCGCGCTCACTGCAAGTGGCGCGCGGCTCGGA 1724
Qy 908 G-----AACTTG-----TAGCTGAAGAGCGAG-----AGTAGCTC 939
Db 1725 GCCACAGGGTTCCCGCACTCGGCGCTTGGCGGA--GGCGAGGCTGTTTACGCAAGAAC 1782
Qy 940 TG-----GCCA--CCAGCTGGGC-----GAGC-----GAGC-----959
Db 1783 CGACCCCGACGACTGCGCTTGGGCGAGCGAGCGGGGTGGCGGAGTGGTGAC 1842
Qy 960 -----TCA-----CCCTGCTC-----CAACCCAGCCCG-----982
Db 1843 TCAGAGGCTCAGGTGCGCTACCCAGCTCACTGCAAGTGGCGCGCGGCTGGCG 1902
Qy 983 CAAGTCTAAGTCTTTC-----AGAGGTG-----1009
Db 1903 CTGGTGTGTGAGAGTCTTGGGCGCTGCTGACCCCGAGCGGACAGAGGTGTCA 1962
Qy 1010 -----GAGGTGTG-----GAAGGAGTGGCTGCTCTCCAACTATGCCAAGC-----GG 1053
Db 1963 GCAGCCAGGTGTGTATACATACGGGT--CT-CTCTCCAGC--CGCCAGCCAGCGCGG 2017
Qy 1054 CGGAGA-----GCTGTCT--TCTGT--CTCCTTGGAGAAAGTCTGTGTCCTG 1102
Db 2018 CGGCGACCGCTGGGCGAGCGCGAGGCTCTCTCTGATGAGC--CTGCGCGCGCGC 2075
Qy 1103 ATTTATGAAGCTA--TAATAGATATATAGTCTTTC-----TACCTTTTTCAGGAAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTACAGGTTTCGGGCGAGCGGTTTGTCCAGAACG 2135
Qy 1156 GTGACTTTCTGTAAATGCGATGTATA--TTAACTTTTATAAAGTT 1203
Db 2136 CGGCTCCCGACCCAGATCGCGTATATAGATATGCAATTTTATTACTT 2186
```

RESULT 14

```
US-10-142-426-543
; Sequence 543, Application US/10142426
; Publication No. US2004004833A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
```

```
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C224
; CURRENT APPLICATION NUMBER: US/10/142,426
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-142-426-543
```

```
Query Match 47.4%; Score 3576.5; DB 12; Length 3721;
Best Local Similarity 38.9%; Pred. No. 6.2e-170;
Matches 785; Conservative 0; Mismatches 389; Indels 843; Gaps 110;
```

```
Qy 4 CTGGG--GTGCTGCCAGGAAAGCAATTCGT--GAAGTTAATG--GTTTGTAGTGATT 57
Db 1252 CTGGGCGCATGCCCTTGATCTCTGGACCTGTAGAGCAGCAGCTACTCGAGTGGCT 1311
Qy 58 -----TTTAAATCCTTG--CT--GGCG 75
Db 1312 GTGGACGTGGGAGCGGCGCCCTGGGGCAACAGACCGTTGTCTTCTGGTCTTGAGCGG 1371
Qy 76 GAGA--GGCC-----GCTC--TC--CCCGTATCAGCGCTTC-----CTC 111
Db 1372 GGGACGCTCTCAAGTTCTCTGTCGGGCCAATGCCAGACCTCAGGACAGCTGTGGGCTC 1431
Qy 112 A-----TTC-----TTTGAATTC-----CGGCTCCGCGT 137
Db 1432 AGTGCTTCTCTGGAGGAGTTTGAGACCTACCGCGCGGACAGGTGTGAGCGCGCGCT 1491
Qy 138 -----CTTCGGGCTCAGAC--AGCGCGA-----GGAAGCTGTTT 171
Db 1492 GCGGACAGAGGCGCGCTGCTGAGCTTGAGAGTGGAGCGAGCTTCGGGGGCGCTGCTG 1551
Qy 172 GCAATTT-----AAGCGGG-----CTGT--GAACGC--CCAG-----199
Db 1552 GCTGCTTCCCGCGCTGGTGTGCGAGTGTGCTGCTGCGTCCAGCAGTACTCGGGG 1611
Qy 200 -----GGCGCGCGG-----GGCGCGCGCGGAGCGG-----225
Db 1612 TGTATGAGAGTGTATCGGAGTTCAGGACCCCTACTCGGGTGGGCGCGCGAGCGCTCC 1671
Qy 226 -----GCCATTTTGAATAAAGAGC--GTGCTTC 253
Db 1672 TGCATCTTCTCAGCCCGGCGACAGAGCGGCTTTGAGCAGGAGCTGTCCGGGCGCAGC 1731
Qy 254 ----CAGGC-----AGGCTCTATAAGTACCGCGCGGCGAGCGTGGCGG 294
Db 1732 ACCTCAGGCTTAGGGAGTGCACAGCACTC--CTGCGGCGCAGCGCTCTCCGAG--GACCGG 1789
Qy 295 CG-----TTGCAAGTCA-----CTGTAGCGAGCTTCTT-----322
Db 1790 CGGGGCTGTGTGCGGTGAACCTTCTGTGTAACCTGTGCGTGGCGGCTTCTGTGGGAG 1849
Qy 323 --TTGGTTTTC--TTTCTCTTTGGG-----GCACCTCTG-----GACTCACTC---361
Db 1850 CCGTGTGTCCGCGCTTACGCTGGGCTGTGTTCTGTTGGGCTCTCGTGTAGGCGCGGAGCTGG 1909
Qy 362 CCGAGCATGAAGCGCTGAG-----CCCGGTGCGC--GGCTCTACGAGCGGCTG-----409
Db 1910 CCGCGCGCAAGGACAGGAGGCGGCTCTCTGCGCACAGGGCGG--GCGAGGCGGCTGCTGAG 1967
```

QY 410 ----TGCTCCCTGTCGGAACGAG--TCTGG--CCATCCCGGGGGCCGAGG-----453
Db 1968 CGTACGCCCTTGGGCGAGCGAGGCGGCGAGGTCCTCCGGGGCGGGGCGGAGCGGTGG 2027
QY 454 ----GAAGGG-----CCCGG-----CAGCTGAGGAGCGCTGA-----482
Db 2028 CGGTGGCGCGGGTTCCCGGAGGCGCTGCT--GGCGCCCTGTGATGACGAGCGGTGG 2085
QY 483 ----GCT--TGTGG-----ACGATGAG-C--ACTGCTACTCCG 516
Db 2086 GCGAGGCGACGCTGCTGAGGGGCGGGCCCGACGACCTGACTCGGGGCTGCTGCCACG 2145
QY 517 C-----CTGGG--GAA-----GTACCGGAGTCCCGAG--546
Db 2146 CCGAGCAGACGCGCTGCGGAGGAGCGCTGCGCCACTCGCGACCCCGCCACGCGC 2205
QY 547 --AGGCACTC-AGCTTAG--CCAGTGGGAATCTACAGCGCTCATGCTACTACTTCT-600
Db 2206 CTGGGCGCGCGCTTGGGACGCGGACCGCCCTGCT-CCGGCCCTCCGCTTCACTCCTC 2264
QY 601 CGACTGCGAGTAGT-CTGGGCGGAGC-----CAGCC-----CCTG-----GAC 638
Db 2265 CTTCTGCTGCTGGGCG 2324
QY 639 CCGCTGATGCGCGCGCGC-----CTTCCGATCCAGACAGC-----CGA-----GC 677
Db 2325 CCCC-GACGCGCGCTTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2383
QY 678 TGCTCCGGAATCTGTCTCTC-----CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Db 2384 TCACCC-----CCAGCGACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2440
QY 718 TTGCACTGACTGCGCGCTGCTGACACCTCCAG--AAGCGAGGTGCTGGCGCGCT--774
Db 2441 TGAACCGACCTGACCG--CCGATGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2497
QY 775 -CTGC-----CTGGACCGCGGAACTCTCTGCGCGAA-----GCGGA-814
Db 2498 GCAGCGTGGAGGACACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2555
QY 815 -----CGGCGGATGGGC-----CCA-----ACTTGGCGCT 842
Db 2556 CACACGTTACACGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2615
QY 843 GCG--CACTGACTT--CACCAATCCCTT-----C-----C 869
Db 2616 GCGGCGAC-AGACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2674
QY 870 CTGGAGACTAACTGGTGC-----CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2674
Db 2675 CCGTGCCTAGCGCGGCG 2734
QY 897 CGAGGACTGTGAATTTGATGCTTAAGAGCGAGCTA-----GCTCTGG--942
Db 2735 CGAGAGACGGTG-CCAGAACGCGG--GGCGCGGCGCACTCCGAGTGGGTGCTCAAGTC 2791
QY 943 -----CCACGAGC-----TGGG-----CGAGCTCAC-----CCTGC 968
Db 2792 CCCCCCGACCCACCGCGGAGTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2851
QY 969 T--CCACCG 1019
Db 2852 TCGCCCTCCCTTACCG 2911
QY 1020 AGGA-----GTGGCTG-----GTGGCTG-----1030
Db 2912 AGGACTTGTATGATTTGAGTTGACCTTATGCGCGTAGTTTGTGTTTTTTTGTGA 2971
QY 1031 -----CTCTCAA-----ACTATGC-----CAGGCGCGCGCG 1057
Db 2972 GTTTTGGTTCTTTTGGCGTTTCTTAACCAATGACAACTCCGTTCTCGGGGTTGGCGCG 3031

QY 1058 AGAGCTG----GTCTTCTGCTCTCC-TTGGAGAAGG-----TTCTGTTG-----1097
Db 3032 AG-GCAGGGAGGCT--TGGACCGCGGTGGGGAATGGGGGCGCCACAGCTGCAGACCTAAG 3088
QY 1098 -----CCCTGATTTATGAACCTCTATAA-----TAGAGTATATAGT--1133
Db 3089 CCCTCCCGCCACCTGCGAAGGTCCTCCCGCAACCCAGGCGCCCTGGCGTGTGGGTGTG 3148
QY 1134 -----TTGTACCTTTT-----ACAGGAAGT-GAC-----TTTCTGT 1167
Db 3149 CCGCGTGTGCGTCCCGCTGCTGTCGTCGAAGGCGCGGAGGTGGCGTGTGTGTGCGGT 3208
QY 1168 AACATGCGATG--TATATTAAACTTTTATAAAAGT 1202
Db 3209 GCGA--GCGAAGGCTGCTGTGGCGCTGTGTCAAGT 3243
RESULT 15
US-10-123-155-543
; Sequence 543, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-543
Query Match 47.4%; Score 3576.5; DB 14; Length 3721;
Best Local Similarity 38.9%; Pred. No. 6.2e-170;
Matches 785; Conservative 0; Mismatches 389; Indels 843; Gaps 110;
QY 4 CTGGG--GTGCTGCCAGGAAAAAGCAATCTCG--GAAAGTTAATG--GTTTTCAGTGAAT 57
Db 1252 CTGGGCGCATGGCCCTCGATCCTCGGACCTGATGAGGACACCGCTGACTCGAGTGGCT 1311
QY 58 -----TTTAATCCTTG--CT--GGCG 75
Db 1312 GTGAGCTGGAGCGCGCCCTCGGGCAACACGACCGTTGTCTCTGGGTTCTGAGGGG 1371
QY 76 GAGA-GGCGCC-----GCCTC-TC--CCCGTATCAGCGCTTC-----CTC 111
Db 1372 GGGACGCTCTCAAGTTCTCTCGTCCGCGCCCAATGCCAGCACCTCAGGAGCGTCTGGGCTC 1431
QY 112 A-----TTC-----TTTGAATCC-----GGGCTCCCGGT 137
Db 1432 AGTGTCTCTCGAGGAGTTTGAGACTACCGGCGCGACAGGTGTGAGCGCGCGCGGT 1491
QY 138 -----CTTCGGCGCTCAGACC-----AGCCGGA-----GGAAGCCTGTTT 171

1492	GGCGAGACAGGGCAGCGGCTGCTGAGCTTGGAGTCGACGAGCTTCGGGGGGCGCTGTGCTG	1551
172	GCAATTT-----AAGCGG-----CTGT-GAACGC--CCAG-----	199
1552	GCTGCTTTCCCGCTGTGGTGTTCGAGTGCCTGTGGTCTGCGAGCTTCCAGAGTACTCTCGGG	1611
200	-----GGCGGGGG-----GGCGGGGGCGGAGCGG-----	225
1612	TGTATGAGAACTGTATCGGCAGTCAGGACCCCTACTCGGGTGGCCCCCGACGGCTCC	1671
226	-----GCCATTTGAATAAAGAGC--GTGCCTTC	253
1672	TGCATCTTCTCAGCCCGGCCACAGAGCGGCTTTGACGAGACGTGTCCGGGGCCAGC	1731
254	-----CAGGC-----AGCCTCTAAGTGACCGCGCGGAGCGTGCAGCG	294
1732	ACCTCAGGCTTAGGGACTGCGACAGACTC-CTGCGGGCCAGCCTCTCCGAG-GACCGCG	1789
295	CG-----TTGCAGGTCA-----CTGTAGCGGACTCTCT	322
1790	CGGGCGTGTTCGGTGAACCTGCTGTTAACTGTTCGGTGGCGGCTTCGTGGTGGGAG	1849
323	--TTGGTTTC--TTTCTCTTTGG-----GCACCTCTG-----GACTCACATC-----	361
1850	CGGTGTGTCCGCTTCAGGTTGGCTGTGTTCTGGGCTTCGTGAGCGCGGAGCTGG	1909
362	CCAGCATGAAGCGGTGAG-----CCCGGTGGC--GGTGTCTACGAGCGGTG--	409
1910	CCCGCGCAAGACAAGAGAGGCCATCTGTCGCACCGGGCGG--GGAGGCGGTGCTGAG	1967
410	---TGTGCTCTCCGACGCGAG--TCTGG--CCATCGCCCGGGCGCGAGG-----	453
1968	CGTCAGCGGCTGGCGAGCGCAGGGCGCAGGTTCCGGGGCGCGGGCGGAGCGGTGG	2027
454	-----GAAGGG-----CCCGG-----CAGCTGAGGAGCGGTGA-----	482
2028	CGGTGGCGCGGGGTTCCCCCGAGGGCCCTGTCT--GGCGGCCCTGATGCAGAACGGCTGG	2085
483	-----GCT--TGCTGG-----ACGACATGAC--C--ACTGTACTCCCG	516
2086	GCCAAGGCCACGTGTCTGACGGCGGGCCCAAGCTTGACTGGGGTGTCTGCCACG	2145
517	C-----CTGCGG--GAA-----CTG-----GTACCGGAGTCCCAG--	546
2146	CCCGAGCAGACCGCTGCGCAGAGACGCTGCCCACTCCGACCGCGCACCCCCACGCC	2205
547	--AGGCATC-AGCTTAG--CCAGTGGAAATCCTACAGCGCTCATCGACTACATCT-	600
2206	CTGGGCCCCCGGCTGGGACACGCGCACCCCTGTCT-CCCGGCTTCCGCTTCATCTC	2264
601	CGACTCGAGTAGT-CCTGCGCGAGC-----CAGCC-----CCTG-----GAC	638
2265	CCTCTGCTGTGGCGCCGCGCGGGCCCCCGAGCAGCCCCCGCGCTTGGGAGCGCAC	2324
639	CCCTGTGATGGCCCCAC--CTTCCCATCCAGACAGC-----CGA-----GC	677
2325	CCCC-GACGGCGGCTCTATGCTCCCGGCCCGCGCGCTCCACGGGACTTCCCGC	2383
678	TGCTCCGGAATTGTCTATCTC-----CAACGACAAAAAGGAGCTT	717
2384	TCACCCC---CCAGCCAGCCCCGACCGCGGGGTGGTGTCCGCGCCACCGGCCCT	2440
718	TTGCCACTGACTCGCGCGTGTCTGACACTTCCAG-AACGAGAGTGTGCGGCCCGT--	774
2441	TGGACCCAGCCTCAGCGG-----CCCATGTGCTCCGCGGCCCTGTGAGCCCGCCCGCAGG	2497
775	CTCG-----CTGGGACCCGGGAACCTCTCTCCCGGAA-----GCCGGA--	814
2498	GCAGCTGAGAGGCCACTGGGCCCCCCACG--CCCTTCGGGCCCGCACCTGTGCGCGCAC	2555
815	-----CGSACGGATGGGCC-----CCA-----ACTTGCCTCT	842

Search completed: May 7, 2004, 15:16:57
Job time : 96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 7, 2004, 15:04:14 ; Search time 27 Seconds
(without alignments)
4285.867 Million cell updates/sec

Title: X69111
Perfect score: 7544
Sequence: 1 GATCTGGGCTGCTGCCAGGA.....TTAAACTTTTATATAAAGTT 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: Pir1: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1074.5	14.2	1489	D70807	hypothetical glyci
2	971.5	12.9	1079	B70807	hypothetical glyci
3	971.5	12.9	2639	T31328	fibroin - Chinese
4	970.5	12.9	1901	F70806	hypothetical glyci
5	930	12.3	2395	S50820	surface protein ty
6	898.5	11.9	2718	A23475	G surface protein
7	884.5	11.7	1766	A42125	trophozoite cystei
8	860.5	11.4	1574	T13954	MEGF6 protein - ra
9	860	11.4	1381	E70806	hypothetical glyci
10	857	11.4	2704	S09118	G surface protein
11	852	11.3	1306	A70834	hypothetical glyci
12	800.5	10.6	1660	A70869	hypothetical glyci
13	797	10.6	1538	H70846	hypothetical glyci
14	789	10.5	1329	E70917	hypothetical glyci
15	784.5	10.4	914	H70987	hypothetical glyci
16	772	10.2	749	A70812	mucin, submaxillar
17	739.5	9.8	13288	T03099	hypothetical glyci
18	717	9.5	882	E70812	hypothetical glyci
19	714.5	9.5	957	D70835	hypothetical glyci
20	706.5	9.4	853	A70896	hypothetical glyci
21	687.5	9.1	837	E70835	hypothetical glyci
22	684.5	9.1	2352	T30201	Notch homolog prot
23	676.5	9.0	2531	S18188	Notch protein homo
24	675	8.9	1620	T27283	hypothetical prote
25	675	8.9	2555	A40043	hypothetical prote
26	670.5	8.9	5376	T42215	zonadhesin - mouse
27	669.5	8.9	2533	T28674	alpha-SID-immobili
28	668	8.9	778	F70963	hypothetical glyci
29	666	8.8	481	A35628	loricrin - mouse

ALIGNMENTS

RESULT 1

D70807
hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genom
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1489 <COL>
A:Cross-references: GB:AL020202; GB:AL123456; NID:93261554; PIDN:CAAL7751.1; PID:G2924
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3514
C:Superfamily: collagen alpha 1(IV) chain

Query Match	14.2%	Score	1074.5	DB 2	Length	1489;
Best Local Similarity	33.7%	Pred. No.	3.5e-42;			
Matches	432	Conservative	19	Mismatches	561	Indels 271; Gaps 54;
QY	1	GATCTGGGTGCTGCCAGGAAAGCAAAATCTTGA--AGTTAATGGTTTTG--AGTGAT	56			
DB	208	GGLTWGGGAGGVGGAGGCTGGAGRAELFGAGGAGAGTDCGPGATGCTGGHGGVGD	267			
QY	57	TTTTAAATCCTTGTCTGGCGAGAGGCCCGCTCTCCCGGTATCAGCGCTTCTCTCTCT	116			
DB	268	GGWLPAGGAGGAGGGGAGGAGSDG-----GALGGTGGTG-----	303			
QY	117	TTGAATCCGGGCTCCCGGCTTCGCGCTCAGACCGGAGGAGCCCTGTTTGAAT	176			
DB	304	TGAGAGGAGGALLLAGGQ-----GGLGAGGQ-----GGTGGAGG--DGLVGGVGT	350			
QY	177	TTAAGCGG--GCTGTGAACCCAGGCGCGGGCGGGCCGCGGCGGCGGCGCATTTTG	234			
DB	351	GGKGGVGVAGLGAGGAAGQLFSAAGAGNAGVGGAGGGGGGAGGAGADADAPGATG	410			
QY	235	AATAAGAGCGGTGCTTCACGAGGAGCTCTATAGTGACCGCGCGCGGAGCGTGG--	292			
DB	411	GTGFAGGAGGAGG-----AGSSG-----AGTTGGGAGGCGGAGGAGADNP	456			
QY	293	CGCGTTGAGTCACTGTAGCGGACTTCTTTTGGTTTTTCTTTCTTTCTTTGGGGCACCTCTG	352			
DB	457	TGIGTGSDGDTGGAAGAGGAGGA-----AGTGGTGMIGTTGNAGVGGAGGGGGGAG	511			
QY	353	GACTCACTCCCCAGCA-TGAAGGCGCTGAGCCCCGCT-CCGCGGCTGTCTACGAGCGG--	407			

notch-1 protein -
notch3 protein - h
Balbiani ring 3 pr
alpha-SID immobili
hypothetical glyci
notch 3 protein -
notch 3 protein -
hypothetical glyci
hypothetical glyci
notch homolog - se
hypothetical glyci
major ampullate fi
surface antigen - P
transmembrane prot
Xotch protein - Af
period clock prote
surface protein 51

30 661.5 8.8 2531 2 A46019
31 656 8.7 2321 2 S78549
32 654.5 8.7 1700 2 S08167
33 644.5 8.5 2533 2 T28675
34 638.5 8.5 801 2 F70824
35 637.5 8.5 2318 2 S45306
36 635.5 8.4 783 2 E70824
37 634.5 8.4 767 2 E70895
38 631 8.4 2531 2 T31070
39 625 8.3 731 2 C70974
40 618.5 8.2 718 2 A36068
41 618.5 8.2 2543 2 T31687
42 616.5 8.2 2437 2 S42612
43 613.5 8.1 2524 2 A35844
44 603.5 8.0 713 1 UMMS
45 601.5 8.0 2233 2 T28669

QY 775 CTGCTGGGACCCGGAACCTCTCTGCGGAGCCGACGAGGATGGCCCAAC 834
Db 760 GTGGGGTGGGNGGNGADNTAN-MTAQAGGDDG--NGDGGFGGAGAGGG-- 809
QY 835 TTGCGCCCTGCCACTTGAATTCACCAATCCCTTCTCTGGAGACTAAACCT-- 884
Db 810 -----GLTAGANGTGGGGAGG-----DSGNGAIGHGRLTDDPGNGGGTG 850
QY 885 --GGTGTCTAGGAGCGAGGACTGTGA---ACTTGTAGCTGAAGAGCGAGTAGCT 938
Db 851 GNGGTGTG--GGAGIGSLGGGTGGGNGGNGGTGGGEGVGGAGGTGAAGNGDGG-- 906
QY 939 CTGCGCACCGAGCTGGGCGACGTCAACCTGTCTCCACCCACCCCAAGTCTTAAGTCTT 998
Db 907 TGG-----TGGDGG-----GAGGTGT 923
QY 999 TTCAGAG-----CGTGGAGGT-GTGGAGAGGAGTGGTCTCTCCAAACTATGCCAAGCGG 1053
Db 924 GGTGGLDPRVGGSGDGGTGGSGGAGNGGNGNAG-----AGNGNGG 968
QY 1054 CGGCAGAGCTGGTCTCTGCTCTCTTGGAGAGAGGTTCTTGTCCCTGATTTATGAAC 1113
Db 969 TGAAGGTGGT-----NGDAPGVPYPGAGGAGGTGGKGG---TGNNGSGTG--- 1016
QY 1114 CTATAATAGATATATAGTTTGTACCTTTTACAGGA-----AGTGAATTTCTG 1166
Db 1017 -----SGGTGGDGTGGGNGGTGWNKGDTGSGGAGDGGKAPAGGTG----- 1062
QY 1167 TAACAATGCGATGA 1181
Db 1063 -----GAGDGGGA 1070

RESULT 3
T31328
fibroin - Chinese oak silkworm
C;Species: Antherea pernyi (Chinese oak silkworm)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31328
R;Sezutsu, H.; Tamura, T.; Yukuhiro, K.
A;Description: Characterization of the full length fibroin gene of a wild silkworm, Anth
A;Reference number: 220995
A;Accession: T31328
A;Status: preliminary; translated from GB/EMBL/DBJ.
A;Molecule type: DNA
A;Residues: 1-2639 <SEZ>
A;Cross-references: EMBL:AF083334; NID:G3450882; PID:G3450883; PIDN:AAC32606.1
C;Genetics:
A;Introns: 14/3

Query Match 12.9%; Score 971.5; DB 2; Length 2639;
Best Local Similarity 25.1%; Pred. No. 2.2e-37;
Matches 386; Conservative 60; Mismatches 707; Indels 385; Gaps 43;
QY 2 ATCTGGGGTGTGTCAGGAAAGCAAT-----CTGGAAGTTAATGTTTGA- 51
Db 344 AASGAGRGDGGYGGSSSNAANAARRAGHRAAGSNAANAASGAG 403
QY 52 -----GTGATTTTAATCTTCTGTCGCGGAGAG-----GCCCGC 86
Db 404 GSGGGYCWGDGYSNAANAASGAGGAGGAGGYSYGDGSDSNAANA 463
QY 87 CTCCTCCCGGTATCAGCGCTTCTCTTCTTGA-ATCCGCGGCTCCGCGTCTTCGCG 145
Db 464 AANAAGSGGGYGGYSDSNAANAASGAGGAGGAGGYSYGDGSDSNA 523
QY 146 TCAGACCGCGGAGGAGCTGTTGCAATTAAGC-----GGCTGTGAACGCC 197
Db 524 AAAAAAGSAGGAGGAGGYSNAANAARRAGHRAAGSNAANAASGAG 583
QY 198 AGGCGCGCGGGCGGGC-----CGAGGCGG-----C 227

584 AASG-AGSGGGYCWGDGYSNAANAASGAGGAGGYSYGDGSDS 642
QY 228 CATTTTGAATAAGAGGCGTCTTCCAGGAGGCTCTTAAGTGAACCGCCGCGGAGC 287
Db 643 AAAAAAASGARGSG-----YGGYSDSNAANAASGAGGAGGYSYGDGSDS 694
QY 288 GTGCGCGGTTTGCAGGTCACTGTAGCG-----GACTTCT----- 321
Db 695 GWDG-GYSDSNAANAASGAGGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 753
QY 322 TTTGTTTCTTCTTCTTGGGCACTCTGCACTCTCCAGCATGAAGGCGCTGAG 381
Db 754 SAASNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 813
QY 382 CCCGTGCGCG-----CTGCTACGAGCGGTGTGCTGCTCTCGAACGCGAG 429
Db 814 ---GGYCWGDGYSNAANAASGARGSGYGGYSDSNAANAASGAGGAGGYSYGDGSDS 870
QY 430 TCTGCCATCCCGGGCGAGGAGGCGCGAGCTGAGAGCGCTGAGCGCTTGTCT 489
Db 871 SGAGVGGYCWGDGYSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 917
QY 490 GAGCGACATGAACA-----CTGCTACTCCCGCTCGCGGAACCTGTACCCGAGTC 541
Db 918 GSSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 977
QY 542 CCGAGGAGCACTAGCTTTAGCCAG-----GTGG-----AAATCTTACAGC-CCGTCTAT 588
Db 978 SANAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1037
QY 589 CGACTACATTTCTGACCTGC-AGGTAGTCTCTG-----GCCGAGCGAGCCCTTGGACCCCT 643
Db 1038 YGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1091
QY 644 GATG-----GCCCGCACCTTC-----CCATCCAGACAGCCGAGCTCGC 681
Db 1092 GAGRGDGGYGGSSNAANAARRAGHRAAGSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1151
QY 682 TCCGGAATTTCTATCTCTCAACGACAAAGAG-----CTTTTCCACTGAC 728
Db 1152 YWGDGYSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1211
QY 729 TCGGCGGTGCTCTG---ACACCTCCAGAACGAGTGTGCGCGCGCTTCTGCTCGGAC 785
Db 1212 GNGGYCWGDGYSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1270
QY 786 CCCGGAACCTCTCTCTGCGG-----AAGCCGAGCGGAGGAGTGG-GCC 829
Db 1271 AAGSGAG-----GAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1322
QY 830 CCAACTTCGCGCTGCCACTTGACTTCAACCAATCTCTTCTCTGAG----- 875
Db 1323 SNAANAARRAGHRAAGSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1382
QY 876 -----ACTAACTCTG-----CTCAGGAGCGAAG----- 902
Db 1383 AAAAAAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1442
QY 903 -ACTGTGACTTGTAGCTTGAAGCCAGAGCTAGCTCTGCCACCACTCTGCGGAGCTC 961
Db 1443 YGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1500
QY 962 ACCCTGTCTCCACCCACCTCTTCTAAG-----TCTTTTCTAGA 1004
Db 1501 -----GYGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1555
QY 1005 GCGTGGAGGTGGA-----AGGAGCTG 1026
Db 1556 GAGGAGGYSYGDGSDSNAANAARRAGHRAAGSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDSNAANAASGAGGAGGYSYGDGSDS 1615
QY 1027 ---GCTGTCTCTCAACTATGCCAAG-GCGCGGAGAGCTGCTTCTGCTCTCTCTG 1082

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C;Genetics:
A;Genetic code: GGCs
A;Introns: 472/3; 1310/3; 1821/3
C;Superfamily: G surface protein

Query Match 12.3%; Score 930; DB 1; Length 2395;
Best Local Similarity 23.4%; Pred. No. 1.6e-35;
Matches 481; Conservative 67; Mismatches 637; Indels 874; Gaps 111;

QY	3	TCCTGGGTG-CTGCGAGAAAGCAAT-----TCTGGAAGTAATGGTTTGGT	53
DB	248	TCKSFRDGTCTTKANGCGVTRTTCAATTOAACVKNSSNGDCYWTGTACVCKTCTNA--	305
QY	54	GATTTTAAATCCCTGTGCT-----GGCGGAGA-----GGCGCGCCTC---TC	91
DB	306	PATWTNSACAGFVTGCTTKSGGCGVSNAGCAANVOACVKNSTGCTGCTDCTCKEKC	365
QY	92	C-----CCGGATC-----AGGCTTCTCATCTCTTGA-----ATCC---GC	126
DB	366	ANAPTNNHDLCTSYLSTCVKSGGCGQPRTCANAPTITINDACEYLPANNCLTKSG	425
QY	127	GGC-----TCC-----GGGTCTTCGGGCTCAGACCGCGGAGAACCTGTTT	171
DB	426	GGCVPTNTCAVITLEACVKNSSGATCFWDSASSTCKDKTCAN-APSTNTDLCVAFLS	484
QY	172	GCAATTTAAGC-----GGGC-----TGTGAC--	193
DB	485	SCTVNSTNAGCVDKTCENSLAQTICDKDLANKACIWKGCYKRECVLASSTTATHADCQT	544
QY	194	--GCCCA--GGGC-----CGG-----CG-GGGCGGGGC--	217
DB	545	YDVCTLSNTGCTGVPLPLKCEAITIEACNIRLOVTSVKSYPGCGWSSCIDKACST	604
QY	218	-----CGA--GGCGG-----GC-----CATTTTG-----	234
DB	605	APKTTATTSDDGTYSKGCVANPNVNGSIQCGDLPTTCAARKSTENCEIIRTPPTCLMN	664
QY	235	AATAAGAGGGCGTGCCTT-----CCAGGCAGGCTCTATAAGTACCGCGCGG	282
DB	665	SATSACVEKSCSTASVTTTGTFLVFNSTNCLAVLSNACIANNTADGCIPEKSSCGSLT	724
QY	283	CGAGCGTGC-----CGCGGTGACGGTC---ACTGTAGCGGACTTCTTTTGGTTTCTTTC	335
DB	725	TSANGVGSKANGDCYNGTTCVDRCTSNISLSLTHSGCNGVLTNC--TVNNAKTIQSLA	782
QY	336	TCTTTGGGCAC-----CTCTGGAC---TC-----ACTC-----CCCA	365
DB	783	TACTSYGSENCKOTSAGKCIWGTACRNATCLDASDTNSFSDGECSEAYPTPSDCTV	842
QY	366	GCATGAAGCGCTGAGC-----CCGGTGCGGGCTGTAGC-----AGCGG---	406
DB	843	LYKTGAV-GCVTRGANGSDYVSQAQCVRTLTAASDDCTWKSSHKCFPSNTYLSGACSTFQ	901
QY	407	GTGTGC-----TGTGCG--GAA--CGCAG--TCT-----GGCCAT	438
DB	902	GTKLCEAIKTCGNNAVAVETDCTSCAVVTGTSLTHAGCQGYSTICTANGSGTACLT	961
QY	439	CGCCC-----GGGGC-----CGAGGGAAGGGC-----CCGCGACCTGAGGACCGCT	480
DB	962	LAAACSTYTAQDCVRRDGLSCFNWAGASPAACQDITSTNCNALASLTGATHASCOAYS	1021
QY	481	GAGCTTGTCTGAC-----GACATGACACAC-----TGCTACTCCCG-CTTGGG	523
DB	1022	TA-CTSIISDGTACQAIQAACSSYSALTACTKRTDGLKCFWNTGATPAAACQDITSTNCNA	1080
QY	524	GAACGTGGTACCGGAGTCCCGAGAG-GCACTCAGCTT-----AGC-----CAGGTGGA	570
DB	1081	LASLTGATHATCOAVSTACTSISDGTACQALQAACSSYSALTACTQRTDGLKCFWNTGAT	1140
QY	571	AATCCTACAGCGCG-----TCATCGACTACATCT-----CGACCTGAGGTAGTCTCT	618
DB	1141	PAAACQDITSNALASLTGATHASCOAVSTACTS-SDGTACQALQAACSSYSQAQTACT	1200

QY	619	GGCGGAGC-----CAGCCCT-----GGAC--	638
DB	1201	ORTDLKCFWNTAATPAAACODITSAAGSLITGLSSATHATCOAYSTACYNIAAGACQA	1260
QY	639	---CCCTGATGGCCGCCACCTCC-----CAT---CCAG-----ACAGCC---	673
DB	1261	FSTCEALTGSLNLTWITICAFSTTCSVKRDGTGCVTIQASCTGYTIANCYSTAGCTAN	1320
QY	674	-----GAGCTC-----GCTCCGGAACCTGTTCATCT-----	698
DB	1321	SGDSACQAIASSTICEQIKLGSSPAFDDTKCNTFKTGIALGTSGCTKTCANKTTPFAH	1380
QY	699	---CCAGCACA---AAAGGAGCTTTTGCACATGATC-----GGCGGTGCTCT---	741
DB	1381	SDCNTWLTCTSNNAVSPCTTMAATCASLTTCVAVEGECVSGVSTCVKTCDTAS	1440
QY	742	-----GACAC--CTCCAGAACGAGGTGC-----TGG-----	766
DB	1441	ADTSPDFAECIAVLTCTVARTGCGOARATCASYSQCKCFNSTGCKCFWPNPNKTCV	1500
QY	767	---CGCCCGTT-----CTGCCTGGACCCCGGGAACCTCT---CCTGC	803
DB	1501	DINCGNIBATTTYDTHAECVAVDTQLCTVRATNGAAVPGCMARGACSSYSIEQCKTNP	1560
QY	804	CGGAA-----TC-----ACAAATCCCT--TCCTGGAGACTAAAC	825
DB	1561	SGGVCVWNTNLTTPVCQDKSCTTAPTATATHADCDSEYFSTATIKTVVATPDTNGAAVL	1620
QY	826	GGCCCAACTT-----CG-----CCCTGC-----CCACTTG	851
DB	1621	GGCQOATAACSSYIHOEQCFRNATDLCGWNGTQCADKSCATAPATDYDDNDKCRAYFNN	1680
QY	852	ACT-----TC-----ACAAATCCCT--TCCTGGAGACTAAAC	882
DB	1681	KTVASSGGGCVDPIDTCESMTQKCVSDKTRSCYVNGTACITFTCENAPDSFASBDC	1740
QY	883	CTGTGTGCT-----C-----AGGAGCAGAGACTGTGAACCTT-GT-----AGCC	919
DB	1741	NTYLAGCTLDSVKCKTKVCEDFAFATDALCKSALSTCTTNGTNCVTRGTCFQAQNAQCV	1800
QY	920	TGAAGAGC-----CA-----GAGCTAGCTC	939
DB	1801	TSSANVQEWMPAVGTNOAYCTVKTCTAPATLTSACASFFNCTTNGGGCVTKSTC	1860
QY	940	T-----GGCC-----ACC-----AGCTGGGCGAGC-959	959
DB	1861	SAVTIDVACTALNGTVCAWDSAQNKCRDKCQDPFSGTTHAACQTRAGCTAGANGKCAR	1920
QY	960	--TC-----ACCCTGCTCC-----ACCCACCCCA--AGTTCTAAGT	995
DB	1921	VONCBQTTIRACIEGTNGPCLWNNYNSDGSKGACPRYTSCKSLAWNDSITCKWISNQ	1980
QY	996	CTTTTCAGAGC-GTGGAGGTGTGGAAGGAGTGGC-----TGCT---	1032
DB	1981	CITN--GSNCIGITLCSFTNTDGGCVSYDGACIOSVPALNSSDPKVPKPYTSCADAFY	2037
QY	1033	-----CTCCAACT---ATGCCAGGCGC---GGC-----AGAGCTGTCTCTGTCTC	1077
DB	2038	ITTHSCQOTASKCKITNGTGTGIALGACSSYTTQAGCYFNDKGALLTSGAITST--GICTW	2095
QY	1078	CTTGAGAAAGGTCTGTGTGCCCTG--ATTATATGAC---TCTA---TAATAGAGT--	1125
DB	2096	DTTASSCRDQSCADLTGTTTHATCSQLSTCISDGTSLVKGACTSYTTCTTACTTAVGSDG	2155
QY	1126	-----ATAT-----AGTTT-----TGAC-----CTT-TTT	1146
DB	2156	ICYWELASSTNNNTAKRLLACADIQNGTSTNVCAVALSSCVSNGTACIAKANCSTYTK	2215
QY	1147	TAC-----AGGAAGGTG-----ACTTCTGTAAACA---TGCGATGT	1180
DB	2216	TACNSGGLDGCVFTQSTATGAAAGTGTCACTTANSQVCAQAAKDRCSWTAASGT	2275

QY 1181 ATATTAAACTTTTATATAA 1199
 Db 2276 GTTAVASKCAVHTCATNOA 2294

RESULT 6
 A23475
 G surface protein - Paramesidium primaurelia
 C:Species: Paramesidium primaurelia
 C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 07-Dec-1999
 C:Accession: A23475
 R:Prat, A.; Katinka, M.; Caron, F.; Meyer, E.
 J. Mol. Biol. 189, 47-60, 1986
 A:Title: Nucleotide sequence of the Paramesidium primaurelia G surface protein. A huge pro
 A:Reference number: A23475; MUID:87060934; PMID:3783679
 A:Accession: A23475
 A:Molecule type: DNA
 A:Residues: 1-2718 <PRA>
 A:Note: the authors translated the codon TGC for residue 2665 as Trp
 C:Genetics:
 A:Genetic code: SGC5
 C:Superfamily: G surface protein

Query Match 11.9%; Score 898.5; DB 2; Length 2718;
 Best Local Similarity 24.68; Pred. No. 4.4e-34;
 Matches 453; Conservative 56; Mismatches 630; Indels 701; Gaps 102;

QY 3 TCTGGGFGTGCAGGAAAGCAATCTCGAAGTTAATGGT-----TTTGAG 52
 Db 835 TCTVYKVGAGC-----VKSANCSYMTSAQCHKLTNLNTANDCKWIVDRCYALSSFA 890
 QY 53 TGATTTTAAATCC---TTGCTGGCGAGAGCGCCCTCTCCCGGTATCAGCGCT--T 107
 Db 891 TGACTTPTKGTWCEGYRAGCTNTVGAASASCTLDCTLKT-----GSLTAD 939
 QY 108 CCTCATCTTTGAATCCGGCGTCCGGGCTTCGGCGTCAGACGACGCGAGGAGCCT 167
 Db 940 CQALDSTCSVKRGT-----GCATQSTCAGYSTAANCFR---SSASGTAGYC- 985
 QY 168 GTTTGCAATTAAAGC--GGGTGTG-----AAGCCCGCAGGC-----CGCGGGGG 211
 Db 986 AMNTNCQSVTSAECAFVTGLTGLDHSKCOLYHSSCTSLKDGTCQGEYKNTCSGYAATNN 1045
 QY 212 C-----GGGCGCAGGC--GGGCGATTTGAATAAGA--GGC-----GTCC--CTTCCAG 256
 Db 1046 CATSGQKCFDFVECLRFSNCASITGTLTATCGTYDAGCVANVNGTACQEKLATCDLY 1105
 QY 257 GCAGGCTCTAAGTGACC-----GCCGCGGAGCGTGGCGGTTGCAGGTCACT--- 308
 Db 1106 LTQNSCSTSAATAADKCAWSGTACLAVTTVGTHCAVTVTGLTDLICAAYNANCTANKA 1165
 QY 309 GTAGCGGACTCTTTTGGTTTCTTCTCTTGGGCGACCTCTGGAC-----TCATCTCCC 364
 Db 1166 GTA-CQEKATCNLYT--TEATCSTSAATAADK-----CAWSGAACLAVTTVATECAV 1217
 QY 365 AGCATG-----AAGGGCGCTG--AGCCCGGTGGCGGCTGTACGAGGGGTGCTGCC 416
 Db 1218 TG--TGLTDLICAAYNANCTANKAGTACQEKATCNLYTTEATCSTSAATAADK--- 1271
 QY 417 TGTGGACAGCAGTCTGGCCATCGCCGGGCGGAGGAGCGCCGGCAGCTGGAGGAC 476
 Db 1272 --AWSGAACLAVTTVATECAVTVTGLTNAIC-----AAYNANCTANKAGTACQEKAT 1323
 QY 477 CG--CTGAGCTGTCTGGACGACA--TGAACCA-----C-----TGCTACTCCC--GCCT 519
 Db 1324 CNLYTTEATCSTSAATAADKCAWSGAACLAVTTVATECAVTVTGLTNAICAAVYANCT 1383
 QY 520 GCGGAAGCTGTACC-----CGGAGTCCCGAGGAGGACTCAGC-----TTAGCA--G 565
 Db 1384 ANKAGTACQEKATCNLYTTEATC--STSAATAADKCAWSGAACLAVTTVATECAV 1440
 QY 566 GTGGAAATCCTACAGC-----GCCTCATCGACTACAT----- 598

Db 1441 GTGLTATCATYNAGCINLKDGTGCOBAKANKDYTTSNKCTAQTSTLSCLWIDNSCYP 1500
 QY 599 -----CT-----CGACCTCAGGTAGTCTCTG--GGCG--AGCCAGC-----C 631
 Db 1501 VTDLNGSVITGLGFVHAQCOAYSTGCTSVSDGSKQDFKSTCEQVPGTTLGCTKTASTPKC 1560
 QY 632 CCTGGAC-----CCCTGATG-----GC-----CC-----C--CA 654
 Db 1561 YLOGSACIISNVATDCAKITGASGTITFEICQSVNTGCSVNRARSACVQQAQCSGYTS 1620
 QY 655 CWTTC-----CCA-----TCAGACAGCC-----GAG-----CT-----CGTCCCGAA 688
 Db 1621 AMTSCYKSGAGLCIASTNTDTACVAATAATCAATDAVVLGAGNYSSANCMKAGCTNGTTA 1680
 QY 689 CTTGTCTATC-----TCCAACGACAAAGAGAGCTTTTGCCACTGACTC-----GGCG 735
 Db 1681 CVAKTCANAAGITFNHTNCNSYLNTCTVNSGNSAQCTWASKCAQOTQASCLYSVEGCV 1740
 QY 736 TGTCTCT-----GAC-----ACCTCCAGAAC--GCA 758
 Db 1741 VGTSCVRKTCDTAATDTRDDDECTCTYQOCTVARLGACQARAACATYKSLQCKFNTS 1800
 QY 759 GG-----TGCTGGCGCCCGTT-----CTGCTGGGACCCCGGAA 793
 Db 1801 GKKCFWNPNTKCTVDLNCNIEATTLTDHNECVAVDATLACTVRAATNGAAGCMARGA 1860
 QY 794 CCTCT-----CCTGCCGG-----AAGCCGAC-----GGC----- 818
 Db 1861 CASYTIIEOCKTNASNGVCVMTNANLPAPAQDKSCTSAFTSTTHNDYAYNTATVK 1920
 QY 819 -----AGGATGGGCCCAACTT-----CG-----CCTCTC----- 844
 Db 1921 CTVATPNSGGNPTLGGCQTAAACSSVIDKEQCINANGDPGWNQTQCADKSCATASA 1980
 QY 845 -----CCACTTGACT-----TC-----ACCAATCCCT 867
 Db 1981 TADYDDTTCRAVITNKCTVSDSGSCVEIPATCETWTKQCYNKAGDPCTWTGTACIT 2040
 QY 868 TCTGGAGA-CTAAACCTGTGCT-----C-----AGGAGCGAAGGACTGTGAAC 911
 Db 2041 KSCDNAPDATADECENTYLAGCTLNNVKCTKVCEDFAFATDALCKQAICTCTNGTNC 2100
 QY 912 TT-GT-----AGCCTGAGAGC-----CAGAGC-TAGCTCTGCC 944
 Db 2101 VTRGTCFQALSOAGCVTSTNQOCEWIPAVLNASNVITSPAYCTIKNCSTAPITLSE-- 2158
 QY 945 ACCAG-----CT-----GGCGAGCTCACCTGTCTCCACCCACC-----CCCA 985
 Db 2159 AACAGVFTNCTTKNGGGCVTKSTCSAVTIDVACTTALNGTVCAWDSQNKCRDKCQDF 2218
 QY 986 GTT-----CTA-AGGTCTTT-----TCAGAGC--GTGG-----AGGTGTGA 1019
 Db 2219 GTTHAACQORAGCTAGAGKCARVQNCQETSVRAACIEGTNGPCLTWIDKTYQNTDGTNGA 2278
 QY 1020 A-----GGAGT-----GGC----- 1028
 Db 2279 CFPYTSKSLNWNDSCKWISNKNTNGSNCGITLGCSETNTDGGCVTVGDGACIQSVP 2338
 QY 1029 -----TGCT-----CTCCAACCT--ATGCCAGGCGGC--GGC-- 1057
 Db 2339 DLNSDPKCKVPTSCADAFYTHSDQIASKCTNTGTTCIALGSCSSVTVQAGCVFN 2398
 QY 1058 -AGAGCTGTCTCTGTGCTCTCTTGGAGAAAGGTCTCTGTGCCCTGATTATGAAGTCT-- 1114
 Db 2399 DKGLTYSVITST--GICTWDTTSSCRDQSCADLTGTTHATCSSQLSTCTSDTCLL 2456
 QY 1115 -----TATAATAGAGT-----ATAT-----AGGTTT----- 1135
 Db 2457 KGACTSYTTQTACTTAVGSDGACYWELASATNNNTAKCELLTCADIQNGTATNVCVALS 2516
 QY 1136 -----TGTAC-----CTT-TTTTAC-----AGGAAGGTGACT--TTCT-- 1165
 Db 2517 TCVSNGTACIPRANGCSTYTSKVACNSGGIDGICVFQTOSTATGAAAGTGTCTALMTACTVAN 2576

QY 1166 -GTAACAA-----TGCAGTGATATTAATAACTTTTATATA 1199
 Db 2577 NDOTACQARDRCSTWASGTRATAVASKCAHTCATNOA 2616

RESULT 7
 A42125
 trophozoite cysteine-rich surface antigen 170 - Giardia lamblia
 N/Alternate names: CRP170; cysteine-rich surface antigen CRP170
 C/Species: Giardia lamblia
 C/Date: 05-Dec-1998 #sequence revision 05-Dec-1998 #text_change 01-Dec-2000
 C/Accession: A42125; B42125; S00530; S48056
 R/Adam, R.D.; Yang, Y.M.; Nash, T.E.
 Mol. Cell. Biol. 12, 1194-1201, 1992
 A/Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 gene
 A/Reference number: A42125; MUID:92186850; PMID:1545800
 A/Accession: A42125
 A/Molecule type: DNA
 A/Residues: 1-98 <ADA1>
 A/Cross-references: GB:M83937; NID:gl59124
 A/Experimental source: trophozoite
 A/Note: sequence extracted from NCBI backbone (NCBI:89421, NCBIP:89427); this ORF is not
 A/Note: the authors report but do not show 19 tandem repeats of the sequence of residues
 A/Accession: B42125
 A/Molecule type: DNA
 A/Residues: 1269-1766 <ADA2>
 A/Cross-references: GB:M83933; NID:gl59122
 A/Note: sequence extracted from NCBI backbone (NCBI:89424, NCBIP:89431); this ORF is not
 R/Adam, R.D.; Agarwal, A.; Lal, A.A.; de la Cruz, V.F.; McCutchan, T.; Nash, T.E.
 J. Exp. Med. 167, 109-118, 1988
 A/Title: Antigenic variation of a cysteine-rich protein in Giardia lamblia.
 A/Reference number: S00530; MUID:88089405; PMID:3335828
 A/Accession: S00530
 A/Molecule type: DNA
 A/Residues: 1154-1409, 'A', 1411-1420, 'K', 1422-1425, 'R', 1427-1481 <ADA3>
 A/Cross-references: EMBL:X06741; NID:G9355; PID:G929603
 R/Yang, Y.; Adam, R.D.
 Nucleic Acids Res. 22, 2102-2108, 1994
 A/Title: Allele-specific expression of a variant-specific surface protein (VSP) of Giardia
 A/Reference number: S48056; MUID:94301794; PMID:8029018
 A/Accession: S48056
 A/Molecule type: DNA
 A/Residues: 1-56 <YAN>
 A/Cross-references: EMBL:L25059
 A/Experimental source: trophozoites WBA6
 A/Note: the source is designated as Giardia intestinalis
 C/Comment: This translation was produced by PIR staff from information provided by the
 C/Genetics:
 A/Gene: VSPA6
 C/Keywords: surface antigen; tandem repeat

Query Match 11.7%; Score 884.5; DB 2; Length 1766;
 Best Local Similarity 24.2%; Pred. No. 1.5e-33;
 Matches 423; Conservative 33; Mismatches 636; Indels 655; Gaps 99;

QY 2 ATCTGGGGTG--CTGCCAGGAA-----AAAGCAATTCGGAAGTT-----40
 Db 87 APCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVDAKE 146
 QY 41 -----AATGGTTTTGAGTGA-----TTTTTAATCTCT--TGCTGGCG-----GAG 78
 Db 147 CKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSV 206
 QY 79 AGGCGCCGCTCTCCCGGTA--TCAGCGC-----TTCTCATTC-----115
 Db 207 SDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGY 266
 QY 116 ---TTTGAATCCGGGCTCCGGCTCTTCGGC-----GTCAAGACCAAG 154
 Db 267 TDDSVSDAKECKK--NAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSA--VD 323
 QY 155 CCGGAG-----GAAGC--CTGTTTGCATTTTAAGCGGGCTGTGAA-----C 193

Db 324 CQGSAGYTTDDSVSDAKECKKNAPCTACAG--TADKC-TKCDANGAAPYLKKTNPSPDPT 380
 QY 194 GCCCAGGGCGGGCGGGCGGGCCGAGCGGGCCATTTTTGAATAAAGAGCGCTGCTTC 253
 Db 381 GTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPC-----TACAGT-----ADKCTKC 427
 QY 254 CAGCAGGCTCTAAGTGACCGCCCGG-----CGAGC-----GTGC 291
 Db 428 DANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTAD 487
 QY 292 GC--CGGTTGCA-----GGTCA-----CTGTAG-----CGGACTTC 320
 Db 488 KCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTAC 547
 QY 321 TTTTGGTTTCTTCTCTCTTTGG-----GGC-ACCTCTGA-----354
 Db 548 ----AGTADKCT--KCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAK 600
 QY 355 ----CTCACTCCC-----CAGC-ATGAAGCGGTG-----AGCCCGGTGC--CGCG-----393
 Db 601 ECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDD 660
 QY 394 ----CTGTCA-CGAGCGGTGTGTGC-----CTGTGGAA-C-GCA 428
 Db 661 VSDAKECKKNAPCTA--CAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGS 718
 QY 429 GTCTGG-----CCATCGCCCG--GGGC--CGAGGA-----AGGCGCCGCGAG 467
 Db 719 GYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVD 778
 QY 468 CTGAGG-----AGCGGTGAG--CTTGTGGAC-----GAC 496
 Db 779 CQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTC 838
 QY 497 ATGAACCACTG-----CTATCCCGCTCGGGA-ACTGCTACCCGA-----538
 Db 839 VSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNP 895
 QY 539 ----GTCC-----CGAGAG-----GCACCTAGCTTAGCCAGGTGAATCTTAC 578
 Db 896 PSDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCT--ACAG-----TADKCTKC 947
 QY 579 AGCG-----GTCACTGACTATCT-----CGACCTCCAGGTA 613
 Db 948 DANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGT 1006
 QY 614 GTC--C-----TGGCCGAGCCAGCC-----CCTGGACCCC 641
 Db 1007 DKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTA 1066
 QY 642 CTGATGGCCCCCA-----CCTTCCCATCCAGACCCGAG-----676
 Db 1067 CAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGS-----AGYTTDDSVSDAK 1120
 QY 677 ---CTCGCTCCGGA--CT-----TGTCTCTCCACGACAAAAG 711
 Db 1121 ECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDC-----QG 1171
 QY 712 GAGCTTTTG-----CCACTGCTCGGCGGTGCTCTGACACCTCCAGAC-----755
 Db 1172 SAGYTTDDSVSDAKECKKNAPCT--ACAGTADKCT-----KCDANGAAPYLKKTNPSPD 1223
 QY 756 ---CGAGGTGTGGCG-----C--CGTTCTGCTGGAGCCCGGGAAC-----794
 Db 1224 PTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLK 1283
 QY 795 CTCTCTCTCCCGGAAGCCGGA-----CGGAGG---GATGGCCCCAACCTTCG 838
 Db 1284 TNPSDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTAGTQCFCSCDANGER 1343
 QY 839 CCTGCCCCACTTGACTTACCACAAATCCCTCTCTGAGACTAAACCTGCTGCTCAGGAGC- 897

Db 1344 CDQNDVCARCSCTGAPPENGKCPAATPGCHSSCDG-----CTENAMTNQADKCT-----GCK 1394
 QY 898 -----GAAG--GACTGTGAACCTTGT-----GCTGAAGA-----GC--CA 929
 Db 1395 EGRYLKPESIAAGSGTCL-TABECTSDTTHFTYKEKAGDKMCLSCSDATHGITGKKCA 1453
 QY 930 -----GAGCTAGCT-CT-----GGCCA--CCAGCTGGGGAAGCTCACTCTGCTCCCA 973
 Db 1454 LKTLSGAEASTVVCSECTDKRLTPSGNACLNDNCPAGTYADNINGVSVCSCHATCAECNG 1513
 QY 974 CCCACCCCAAGTCTAAGGTCTTTT--CAGA--GCGTGGAGGTGTGAAGGAGTGGC 1028
 Db 1514 DANAASCTACYPGSLLYSGGTAGTCVKECTGAFINCAAGQC-----TANVGAKYC 1566
 QY 1029 TGC-----TCTCCAAACTATGCCAAGCGCGCGCAG-----AGCTG-----GT 1066
 Db 1567 AQCKDGYAPDGICTTVAAGRDASVCTAAD-GKCTKACAGETLMGCGVGVAKLPKSV 1625
 QY 1067 CTTCTGTCTCTCTGGAG-----AAAGTTCTGTGC--CCTGATTATGAACCTCTA 1116
 Db 1626 CTLASNGKCTMCAANNOAPVEKPECESEGCANNDNSNACTECLPGYKAGDKCFKCTA 1685
 QY 1117 TA-----ATAGAGTATAGGTTTGTACCTTTTACAGGAAGGTGACTTTCTGTAA 1169
 Db 1686 SSGNNNQITGVANCVTCAPPAG--NGPVTCTYIKTDGNDTGGSVNKSGLSTGAIAGISV 1743
 QY 1170 CAATGCG 1176
 Db 1744 AVVVVVG 1750

RESULT 8

MEGF6 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
 C:Accession: T13954
 B:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1574 <NAK>
 A:Cross-references: EMBL:AB011532; NID:g3449293; PIDN:BAA32462.1; PID:g3449294
 A:Experimental source: strain Sprague-Dawley; brain
 C:Genetics:
 A:Gene: MEGF6

Query Match 11.4%; Score 860.5; DB 2; Length 1574;
 Best Local Similarity 25.7%; Pred. No. 1.7e-32;
 Matches 364; Conservative 30; Mismatches 552; Indels 473; Gaps 79;
 QY 4 CTGGGGTGC-----TGCAGGAAA--AGCAATCTTGGAGTTAATGTTT 49
 Db 251 CAENG-GCMHICQELRLAHGCHPGYQLAARDKTCEDVDEALGLA----- 297
 QY 50 GAGTGAATTTAAATCTTGTGGCGG-----AGAGGCGCGC----- 86
 Db 298 QCAHGLNTQSFVKVCHAGYELGADGQCYRIEMIIVNSCEAGGCGSHGCSHTSTGPL 357
 QY 87 CTC-----TC-----CC--CGGT-----ATC-----AGCGTTCC 109
 Db 358 CTCPRGYLDEDKTCTCIDDCANSPCCQAACANTPGGYECSFAGYRLNTDGGCEDVD 417
 QY 110 TCAT-----TCTTTGAATCGC-----GGCT----- 130
 Db 418 ECASGHGGEHCNSLNASGFCFCEAGYRLDERRGCTSLSESVVDLDGLRPFVRLPHI 477
 QY 131 -----TCTTTGAATCGC-----GGCT-----CC-----CGGTTCTCGCGTC 147
 Db 478 AVLRLDELPRLFQDDYCAEEAAAEALRGHTLTETKVFCLDHSFGHDCSLTCDRCNGGTC 537

QY 148 -----AGACACGCGGAGGAAGC--CTGTTTG-----CAATTTAAG 181
 Db 538 FPGQDGDCEBWTGIIICNETCPDPTFKNCS9PCTQONGTCDPVLGACRCPFGVSGAH 597
 QY 182 CGGCT-----GTGAACCCCGAGGCGC-----GGCGGGCGGGG-----CGAGGCGGGC 227
 Db 598 CEDCPKPGFYKHKRCKCHCANRGRCHRLYGACLDPLGRFCHLACFPWAFPGGSED 657
 QY 228 --CATTTTGAATAAAGAGGGGTG-----CCTTCAG--GCAGGCTC--TATAAGTGA 273
 Db 658 CLCEQSHTRSCNPKDGCSCSKAGPQGERCOAECESGFFGPGCRHRCTCQCPVACDPVSGE 717
 QY 274 CGCGCGCGGAGGCTGCGCGCTTGAGGTCACTGTAGCGGACTCTTTTGGTTTCTT 333
 Db 718 CRTQPPGYQDEDCGQCEPVTGPNVCS--GSCSVG--APCHRVTEGLCLCPKGTGECGA 775
 QY 334 TCTCTTTGGGCG--ACCTCTGAGCTAC-----TCCCCAGCATGAAGCGCTGAGCCCG 386
 Db 776 DCP--EGRWGLGCBICPACBHA--SCNPETGTCLCLPGFV-----GSRQDT--CSAGW 825
 QY 387 TCGCGGCTGTACGAGCGGTGTCTCTGTCTGCGAAGCAGCTCTG-----CCATCGC 441
 Db 826 YETGQ--IRACANDGHCDPT-----TGRCSCAPGWTGLSCORACDSHGWGPDCHIPCNC 879
 QY 442 CGGCGCGGAGGAGGCGCGGAGCTGAGGAGC--GCTGAGCTTGTCTGAGCAGCATG 499
 Db 880 SAGHNC-----DAVSLCLCE--AGYEGPRCEQSCRQYVYPSCEQK-----RCEHG 926
 QY 500 AACCACTCTACTCCG-----CCTG-----CGGAACTGTGATCCCGAGTCCCGA 545
 Db 927 AACDHVSG--ACTCPAGWRGSCFCHACFAGFFGLDCLDSACNCSAG--APCDAVTGCICPA 983
 QY 546 GAGG--CA-----CTC--AGC--TTAGCCAGGTGAAATCCTACA-----G 580
 Db 984 GHWGPRCAQSPPLTFGLNCSQICTCFNGASCDSTVGQCHCAPGMWGTCLQACPPGLYG 1043
 QY 581 CGCGTCATC--GACTACATCTGACCTGACCTGAGTAGTTC--CTGGCGGAGCGAGCCCT 634
 Db 1044 KNCQSHCLCRNGRCDPLIGQCTCPGWTGLA-----CENECLPGHYAAGCQLNCSLH 1097
 QY 635 GGACCCCTGATGGCGCCCACTTCCCATCCAGACAGCGGAGCTCGTCCCGAACTTGTG 694
 Db 1098 GG-----IDRLTGLCLPAGWTGDKQSSCVST-----FGVHCEHCACRKGASCHVT- 1148
 QY 695 ATCTCCAAAGCAAAAGAGCTTTTGGC-----ACTGACTCGGCGGTGTC--TGACA 745
 Db 1149 -----GACFCPPGWRGPHCEQACPRGWFGEACAQRC-----LCPTNASCHVHTGECR 1195
 QY 746 CCTCCAGAAC--GCAGGTGCTGGCGCCCTTCTGCTGGGACCCCGGAACTCTCTCTGC 803
 Db 1196 CPPGFTGUSCEQACQPGT-----FGKDEHLQCPGETWACDPASGVTC----- 1240
 QY 804 CGGAAGCGGAGCGGAGGATGGG--CCCAACTTTCGCTGCGCCACTTGACTTACCAA 861
 Db 1241 ---AAGYHGTGCLQRCPSGRYGPGEHIC-----KCLNGTCDPATGAC----- 1281
 QY 862 ATCCCTTCTGAGACTAAACCTGTG-----CTCAGGAGCGAGGACT-----GTGAAC 911
 Db 1282 -----YCPAGFLGADCSLACPGQRFPSCAHVACRQGAACDPVSGACICSPGKTVRC 1335
 QY 912 TTGTAGCTTGAAGACGAGAGCTAG--CTCTGGCCACCAG--CTTGGCGAGCT 960
 Db 1336 ---EHGCPQDRFGKGLKACRNGGLCHATINGSCPLGWMGPHCEHACPAGRYGAACL 1392
 QY 961 CACCC-----TG-CTC-----CCACCC-----CACCCCAAGTCTTAAGT 995
 Db 1393 LECFQNNNGSCBPTTGACLCGPGFYGQACEHSCPSGFHGPQRCVCEQQQAPCPVSGQ 1452
 QY 996 CTTTTCAGAGCGTGGAGGTGTGGAAGGAGTGGC-----TGCTCTCCAAAC-----TA 1042
 Db 1453 CL-----CPAGFHGQFCEKGSBSGSGDCLQCCNCHTGVPCDPISGLCLCPFGRTG 1504

QY 1043 TGC---CAAGCGCGGCA-GAGCTGGTCTTCTGCTC 1077
Db 1505 AACDLDCRRGRFG--PGCALRCGCGGADCDPISSQCHC 1541

RESULT 9
E70806
Hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70806
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Connor, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70806
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1381 <Col>
A:Cross-references: GB:A102022; GB:AL123456; NID:g3261554; PIDN:CAAL7744.1; PID:g292444
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3507
C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.4%; Score 860; DB 2; Length 1381;
Best Local Similarity 30.5%; Pred. No. 1.7e-32;
Matches 418; Conservative 23; Mismatches 614; Indels 314; Gaps 54;

QY 1 GATCTGGGCTGTCACAGAAAAGCAAAATCTGAAGTTAATGCTTTTGAAGTGAATTTT 60
Db 147 GAGCGGNGSAGLWNGGAGGAG-----GSGAAGGNGGWLFGAGTGGIGTG 200

QY 61 AAATCTTGTGCGGAGAG---GGCCCGCTCTCCCGATATACCGTCTCTCTCTCTCT 117
Db 201 APG---AMGTTGNGGNGALLTG-----GGLGAGGNGGTTGGTGGTGGNG 244

QY 118 TGAATCCGCGCTCGCGCTCTTCGGCTCA-GACCAGCCGAGGAGCCCTGTTTGA-- 174
Db 245 GNGALLGAGVG--GAGG---TGGTGAAGAGAGAGTGGNGAGGLFMNGSDGAGG 298

QY 175 -----ATTAAAGCG-----GCTGTGAACGCCAGCGCGCGCGCGCGCGCGCG 218
Db 299 QGGDGAAGAAASAGTGGKGGGCGGTGGAGAGPVLFGHGGAGGNGGCGGTGGMG-- 356

QY 219 GAGCGGGCCATT--TGAATAAGAGCGGTGCTTCCAGGAGGCTCTATAAGTGACCG 276
Db 357 GAGGDTTVAAGTGGGTTGAAGAGG--AAGAGALTSGGLAGG---VAGGTG---G 408

QY 277 CCGCGCGAGCGTGGCGCG--GTTGAGGTCACTGTAGCGGACTTTCTTTGCTTTCTTTC 335
Db 409 TGGTGGNGADAADVFGANGDPGFAGGK--GGNGGIGGA-----AVTGG----- 451

QY 336 TCTTTGGGCACCTCTGACTCACTCCAGCATGAGCGCTGAGCCGCTGCGCGCT 395
Db 452 ---VAGDGTGKGTGGA--GGAGNDAGTGNPFGKGGDGLGGAGGAGGAGTNGGHA 507

QY 396 GCTACGA-GCCGCTGTGCTGCTG-----TCGGA-----ACGCACTGCGCCATC 439
Db 508 GNTGDDGDTGGNGGNGTGGVNGADNTLNPDTPGAGCPGAGGAGGAGGAGGPGGTG 567

QY 440 GCCCGGGCCGAGGAGG-----CCCGCAGCTGAGGAGCGCGCTGAGC 484
Db 568 G--TGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 625

QY 485 T-----TGCTGACGACATGAACCACTGCT-----ACTCCCGCTGCGGAACTGG 530
Db 626 TAGSQTGGVGGDGGAGNGGCGGKAGTGNNGFVDEAGTGGAGGNGGNGGAGAGNGG 685

QY 531 TACCCGAGTCCCG--AGAGGCACCTCAGCTTAGCCA-----GGTGGAAATCTCAGCGC 583

Db 686 TCGSGNGDGGAGGIGGAGNGIPGTGTPAGGTGAKGGDGGAGGAGNAGGAGGQ 745
QY 594 GTCATCGACTACATTCT-----CGACCTGCAAGTAGTCTCTGCGCCAGGAGCCCTTGGAC 638
Db 746 GGNAGCGGAGGAGGAGNAVIPGDVGKAPHDAGSGDGGKGGGSGGTGGSGAPIGGG- 804

QY 639 CCCCTGATGGCCCCCACCTTCCATCCAGACAGCGAGCTGCTCCGGAACCTTGTCTCT 698
Db 805 ---AGTGG-----SGHAGKGGAGI-----GAQGTITVPGNG 836

QY 699 CCAACGACAA--AAAGGAGCTTTTGCACCTAGCTCGGCCG-TGTCTGACACCTCCAGAAC 755
Db 837 GNAAGGNGAGAGGNGSGDFGNTTSGASGSGNGGNGAGTAGSG-----GA 885

QY 756 GCAGTGTGCGCGCGCTTCTCTCTGCGGACCCCG-----GGAACCTCTCTCTG 802
Db 886 GTGTGTGTLGSGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 945

QY 803 CCGAAGCGCGAGCGGAG-----CGATGGGCCCC-----AATTCG 839
Db 946 GAPGPAGAPGPTTGNAGSGGIGGGDGGDGGDGGDGGDGGDGGDGGDGGDGGDGG 1004

QY 840 CTTGCCCATCTGACTTCAACAA-----ATCCCTTCTCTGAGAC 877
Db 1005 -SAGDPTGNGGPGTGPSPWAPPPTPTVQGGDGGAGGTGTNANDGTATGKGG- 1062

QY 878 TAAACCTGTCTCAGGAGCGAAGG--ACTGTCAACTTCT-----AGCCTGAAG 924
Db 1063 -----EGVGSIILGPGGNGGTGGNASATGNGVANAGNGKGGDGGGFGAGNGGAGG 1116

QY 925 AGCCAGAGTAGCTCTGGCC--ACAGCTGGCGAGCTCACCTGTCTCCACCCACCC 982
Db 1117 SVTDGAGTAGNGGNGNATNGTITAGQAPAGNGSAGGKG--GDGN----- 1161

QY 983 CAAGTTCTAA-GTCTCTTTTACAGAGGTGAGGTGTGAAGAGTGTCTCTCTCCTCAAACT 1041
Db 1162 IAGATGTAGNGGNGGNGDGNAGTGGSGNG--GNAGGGGANGDGG----- 1209

QY 1042 ATGCCAAGCGCGCGAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1101
Db 1210 -----AGGAGGAGGCGGIDGGPGDGGGNGGNGGTGAGGNGGNGGTGGVSGAAGG 1263

QY 1102 GATTATGAACCTATAATAGAGTATATAGTTTGTACTTTTACAGGAGGTGACT 1161
Db 1264 DGGNGTGGFAGPGGTAGNGSGG-GTGGAGGPGGTGG---DGGNGVIAGGGGTGGNGAS 1319

QY 1162 TT--CTGTAACA-----ATCGATGTAT--ATTAACTTTTATAAAAGT 1202
Db 1320 GAGGAGTGGFAGNGAGNGGTGGASDGDNGNAGSGATGTTGGNGGT 1368

RESULT 10
S09118
G surface protein 168 - Paramesidium primaurelia
C:Species: Paramesidium primaurelia
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 07-Dec-1999
C:Accession: S09118
R: Prat, A.
J. Mol. Biol. 211, 521-535, 1990
A:Title: Conserved sequences flank variable tandem repeats in two alleles of the G sur
A:Reference number: S09118; MUID:90172419; PMID:2308165
A:Accession: S09118
A:Molecule type: DNA
A:Residues: 1-2704 <PRA>
A:Cross-references: EMBL:X52133; NID:g10049; PIDN:CAA36378.1; PID:g578473
C:Genetics:
A:Genetic code: SGCS
C:Superfamily: G surface protein

Query Match 11.4%; Score 857; DB 2; Length 2704;
Best Local Similarity 23.8%; Pred. No. 3.3e-32;
Matches 440; Conservative 59; Mismatches 630; Indels 716; Gaps 98;

QY 3 TCTGGGCTGCTCCAGGAAAGCAAAATCTGGAAGTTAATGCT-----TTTGAG 52
 Db TCTVYKVGAGGC-----VSIKANSYNTSAQCHKILTNLTANDCKWIVDRCYALLSFA 884
 QY 53 TGATTTTAAATCC---TTGCTGGCGGAGAGCGCGCTCTCCCGGTATCAGCGCT--T 107
 Db TGACTTTFKGNKTMCEGYRAGCTNVTGGAASASCTLDCTLKT-----GSLTTFAD 933
 QY 108 CCTCATTTTGAATCCGCGCTCCGCGGTCTTCGGGTCAGACCAGCGAGGAAGCCT 167
 Db 934 COALDSTCSVKOGTGIIVTQSCAGYGSTATNC-----FRSSASGTAGYC- 979
 QY 168 GTTTGCAATTAAACG--GGCTGFG-----AACGCCAGGCG-----CGCGGGGG 211
 Db 980 AMNTNCQSVTSAECAFVTLGLTDLHSHKCOLYHSSCTSLKDGTCQBYKTACSSYATGNT 1039
 QY 212 CGGG--GCC--GAGGC--GGGCCAATTTGAATAAGA--GGC-----GTGC----- 249
 Db 1040 CANSVQKCFDDATDLRFANCASITGTGLTNTICVTYDFGCVANVNGTACQKELATCAA 1099
 QY 250 -----CTTCCAGGCA--GGCTC-----TATAAGTGAACGCGCGCGGAGC-----G 288
 Db 1100 YLTQNSCSTSTAGTCAMSGSACLTVVDANVATECAVITGTGLTNAICAGYNAKCTVNRA 1159
 QY 289 TGC---GGCGTTGAGTCACTGAGC---GGACTTCT-----TTTGGTTTTC 331
 Db 1160 TACQKKEALCATYAAVQATCSQDAGLCAMSGSACLTVVDANVATECPYITGTGLTNAIC 1219
 QY 332 T--TTCTTTTGGGCACCTCTGACTCACTCCCGAGCATGAAGCGCTGAGC----- 382
 Db 1220 AGYNAKCTVNRA-----TACQKKEALCATYAAVQATCSQDAGLCAMSGSACLTVVDANV 1275
 QY 383 ---CGGCTGCG---CGG---CT-----GCTACGAGCGG 407
 Db 1276 ATECPYITGTGLTDAICAGYNAKCTVNRAAGTACQKKEALCATYAAVQATCSQDAGLCAM 1335
 QY 408 TGTGC-----TGCTGTGCGAAGCGAGTGTGCC-----ATCG----- 440
 Db 1336 SGSACLTVVDANVATECPYITGTGLTNAICAGYNAKCTVNRAAGTACQKKEALCATYAAVQ 1395
 QY 441 ---CCGCGGCGCGAGGAAGGCG-----CCGCGAGCTGAG--GAGCGG---CTGAGCT 485
 Db 1396 ATCSQSDAGLCAMSGSA---CLTVVDANVATECAVITGTGLTDAICAGYNAKCTNLKDG 1451
 QY 486 TGC-----TGGACGACATGAACC-----ACTGCTACTC-----CCGC 517
 Db 1452 TCQDEKATCKLYTTONKCTGTTGTLPLSCLFWDNSCSPITDVTCSAIVQSLDHAQCOAY 1511
 QY 518 CTGC---GGGAAGTGGTACCGGAGTCCCGAGAGGCACTAGCTTAGCCAGTGGAAATC 574
 Db 1512 STGCTSVSDGSKQDFKTTCEYAGT-----ALSTKTATSKYLQ-----SNC 1556
 QY 575 CTACAGCGGTCATCGACTACATCTCT---CGACCTGCGAGGTAGTCTCT---GGCCGAGCCAG 629
 Db 1557 ITI-----SNVATDCAKITGSAHTTYEIQSYNTGCVNRAACVQQAQCSGYTSAM 1611
 QY 630 CCCTGAGACCCCTGATGCGCCCGCCCTCCCATCCAGACAGCGAGCT---CGCTCGG 686
 Db 1612 TSCYKSGAGLCIASTNTDTACVAATAATTCDAVYLTG---NYSSANENKAGCTNNGA 1668
 QY 687 AACTTGTATC-----TCCACACAAAGAGAGCTTTGCCACTGACTC-----GSCC 734
 Db 1669 TACVAKTANAVVIFNHNKNGYLNCTVNSGNSACQTMASKADQTOASCLYSVEGCV 1728
 QY 735 GTGTCC-----TGA-----CAC-----CT-----CCAGAACG-----CAGGT 761
 Db 1729 VVGTSCVRKTCDAATDTRDDTSCSAVQSQCTVARLGACQARAACASYSKSLQCKFNT 1788
 QY 762 GCTGG-----CGCCCGTT-----CTGCTCGGACCCCGG 790
 Db 1789 --SGRCFWNPTNKTCDVNLNENIEASTLYDTHNECVVVDATLACTVTRATNGAAVQGCMA 1846

QY 791 GAACCTCT---CTTGCGGG-----AAGCCGGAC-----GGC----- 818
 Db 1847 RGACSSYTBQCKTNASNGVVMNTNANLPAPACQDKSCTSAFTSTTHNDCYAYNTA 1906
 QY 819 -----AGGGATGGGCCCAACT-----CG-----CCTGCG-- 844
 Db 1907 TVKTVVATPSNGNPLGSCQQTACSSVIDKEQCQINANGPCGWNNGTCCADKSCAT 1966
 QY 845 -----CCACTTGACT-----TC-----ACCAATC 864
 Db 1967 APATADYDDTKRAYITNKCTVSQSGCGVEIPATCETMTQKQCYYNKAGDPYCTGTGA 2026
 QY 865 CTTTCTCTCGAGA--CTAAACCTGTTGCT-----C-----AGGAGCGAAGACTGTG 908
 Db 2027 CTTKSCDNAPATATATACNTYLAGCTLDNVKCKTVCEDPAFATDALCKQAICTCTNG 2086
 QY 909 AACTT--GT-----AGCCTGAAGAGC-----CAGAGC--TAGCTCTG 941
 Db 2087 TNCVTRGTFOALSOAGCVTSSTNQCEWIPAVLNASNVTSPAYCTIKNCSTAPITLTS 2146
 QY 942 GCCACGAG---CT---GGCGGACGTCACTGCTGCCACCCACC-----CC 982
 Db 2147 E--AAACAGYFNTCTGNGGCGVTSTGSAVIDVACTTALNGTVCAMDSQAKCKEDKCC 2204
 QY 983 CAAGTT-----CTA--AGGCTTTT-----TCAGAGC--GTGG-----AGGTGT 1016
 Db 2205 DFGTTHAACQAQAGCTTAGASGKCARVQNCQTSVRAACIEGTNGPCLWIDKYQNTDGT 2264
 QY 1017 GGAA-----GGAGT---GGC--- 1028
 Db 2265 KGACPRYTSCSLNWNNDSSCKWISNKTCTGNSNCVGTILCSENTDGGCTVGDGACIQ 2324
 QY 1029 -----TGCT-----CTCCAAACT---ATGCCAAGCGCGCGGAGA 1060
 Db 2325 SVPALNSDPKVKCPYTSADAFYTHSDCQIASSKCTTNGTTCGIALGSCSSV--TAQA 2382
 QY 1061 GC-----TGGTCTTCTGCTCTCTTGAGAAAGTCTCTGTCCTCCCTGATTATGAA 1111
 Db 2383 GCFYNDKGLTYTSGVITST--GCTFWDTSSSRDQSCADLTGTHATCSSLSTCTSDG 2440
 QY 1112 CTC-----TATAATAGAGT-----ATAT-----AGGTTT--- 1135
 Db 2441 TTCLLKAGCTSYTTTACTTAVGSDGACVWELASATNNNTAKRLLTTCADIQNGTATNVC 2500
 QY 1136 -----TGAC-----CTTTT-----TTACAGGAGGTGACT--TT 1163
 Db 2501 SVALSTCVSNGTACIPKANCSTYTSKIACNSGGLDGICVFTQSTATGAAGTGCALMTA 2560
 QY 1164 CT-----GTAACAA-----TGCGATGTATATAAACTTTTATAAA 1199
 Db 2561 CTVANNDQTACQAARDRCSTWTAASGTGATAVASKATHTCATNOA 2605

RESULT 11
 A70934
 hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70934
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genom
 A: Reference number: A70500; MUID:98295987; PMID:9634230
 A: Accession: A70934
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-1306 <COL>
 A: Cross-references: GB:AL021942; GB:AL123456; NID:G3242298; PIDN:CAAL7449.1; PID:G2909
 A: Experimental source: strain H37Rv
 C: Genetics:

A:Gene: Rv0578c
C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.3%; Score 852; DB 2; Length 1306;
Best Local Similarity 30.8%; Pred. No. 3,7e-32;
Matches 404; Conservative 24; Mismatches 507; Indels 378; Gaps 60;

QY 1 GATCTGGGGTGTCTGCAGGAAAGCAATCTCGAAGTAATGTTTGGTGGTATTTT 60
DB 132 GGILYNGNGNGSG--APGORGAGGAAGLIGNGNGG---AGGVGTTG---GAGGGG 182
QY 61 AAATCTTGTCTGGGAGAGCCCGCTCTCCCGCTATCAGCGCTTCTCATTCTTGA 120
DB 183 AGWLYNGAGAGFGAGAVGNGG-----AGGTAGLFGVG-----CA 220
QY 121 ATCCGGCGTCCCGGT-CTTCGGCTCAGACAGCCGAGGAAGCTGTTGCAATTA 179
DB 221 G---GAGGNGIAGTGTSTAGTGGTAGGAGGAGGAGGAGGAGGAGGAGGAGG 269
QY 180 AGCGGCTGTGAAGCCCGG-----CCCGCGGCGGCGGCGGCGGCGCAT 231
DB 270 NGGAGGGGGGAGAGAGAHATNLGADQAGNGNGAGAGTG--GVGGPFGG--HGLL 326
QY 232 TTGAATAAGAGCGTCTCCAGGCA--GGCTCTAT-----AAGTACCGCCCGCGC 283
DB 327 GLGSGHAGAGGSGGD-----GGAPGNGGATCTWGNLGGTGGNGNPGAGGA 379
QY 284 GAGCGTGGCGC-CTTCAGGTCACTAGCGGACTTCTTTGCTTTCTTTCTTTCT 342
DB 380 GGAGGASVGGSAHANGAPGTTTSGNGGDDGGKAD-ALSSQGTGANGRGDDGGVGN 438
QY 343 GCACCTCTGACTCACTCCCGAGC-----ATGAAGCGCTGAGCCCGGTGCGCGCTGC 397
DB 439 GGAGGAGRGGA-----GGLFGSEAPRPGGAGGTGGAGNGGTQAGDGTGG- 487
QY 398 TACAGAGCGGTG-----TGCTGCTGTGCGAACGAGTCTGCCCATCGCCCGG 446
DB 488 -AGGAGGDSGGGAGSIFGNASAPGAAGSPGNGNGGPGGAGG-EGGAGGLALAAAGQN 545
QY 447 GCCGAGG-----GAAG-----GGCCCGGAGCTGAGGAGCGGCTGA 482
DB 546 GSQAGGDDGAGNGGTPGNGHGAAGALGVNGVGG--AGHGDDPVGAG--GGGS 601
QY 483 GCTTGTCTG-----GACGATGAACCACTGTACTTCCCGCTCGGNACT 528
DB 602 GSTPGANGAPNTPTSGNGNGRGADATG---FGQTGASGGSGGDLVNGAGGAG 658
QY 529 G-----GTACCGGAGTCCCGAGAGGCACTCAGCTTACCGAGGTGGAATCTCT 576
DB 659 GNGSKLPGLRLGNPLDGTGG---NGHAGSGGAWNGGTG--GAGGTGGVGTGS 714
QY 577 ACAGC-----GCGTCATCGACTACATCTCCACTGAGGTAGTCTCT 618
DB 715 GSDGVNGSSAGADHPGTGGTGGTGGKGGDGGGAAPNGVAGSQGPGGAGSDGTGGV 774
QY 619 GSCCGAGCCGCTGACCCCTGATGGCCCGCCACTTCCCATCCAGACAGCGGAGCT 678
DB 775 GNGGGRG-IDGADGATAG-----RGQGG-----AGAGGKGRGT 811
QY 679 CGCTCGGA--ACTTGTCTATCCAAACGACAAAGAGGACTTTGCCACTGACTCGGCGCT 736
DB 812 GG---PGGAGPAGTTG-----SOGAGNGSGGTGDPDGGNGANGSVFT 854
QY 737 GTCCTGACACTCCAGACGAG--GTGCTGGCGCCCTCT-----GCTCGGA 784
DB 855 NNGIG-----NGNGNGNAGPSGAGSGGAGTGTGATSSSIHNGNGNGNGND 906
QY 785 CCCCAGGAACTCTCTGCGGAGCCCG-----ACGGCAGGATGGGCCCCCACTTCGCC 840
DB 907 HALSNGAA-----GNGNGNGNSLRSGGAGGNGNGN-----942
QY 841 CTGCCCACTTGACTTCAACCAATCTCTCTGAGACTAAACCTGGTCTCAGAGCGAA 900

943 -----ASRG-----MGDDGG-----TGAGAGN-AGQINGGA 968
QY 901 GGACTGTGAACCTGTAGCTGAAGAGCCAGAGTAGCTTGGCCACCAGCTG-CGCGACG 959
DB 969 GGN-GSDGCTGSDNPAITGSGRGGDGGVGGGGVAGDAGDGGAGGAGTGGTGLRG 1027
QY 960 TCACCTGTCTCCCAACCCCAAGTCTTAAGTCTTTTTCAGAGCGTGGAGGTGGA 1019
DB 1028 T-----TGAT-----CATGTFDAGADG-----HGNNGTGGVG--GTGGA 1060
QY 1020 AGGAGTGTCTCTCCAACTATGCCAAGCGCGGCGAGAGCTGTCTTCTGCTCTCCT 1079
DB 1061 GGGGNGGAGG-----KALSPTGNGSQAGDGGAGGAGGTGG-----TGG----- 1102
QY 1080 TGGAGA-----AAGGTTCTGTTGCCCTGATTTATGAACCTCTATAATAGAGTA- 1126
DB 1103 DGGCAHGTLFSSLAGTGTGNGGTGTG--GTGGAGGAGTGTGATGATCAAGRAG 1160
QY 1127 TATAGTGTGTACTTTTACAGGAAGGTGACTTTCTGTAACAATCCGATG 1179
DB 1161 NGVGGSGGLGSAFPGGT-----GGMGGAG-----GTSTVSAGDGGRG 1200

RESULT 12
A:Gene: Rv2490c
C:Superfamily: collagen alpha 1(IV) chain

Query Match 10.6%; Score 800.5; DB 2; Length 1660;
Best Local Similarity 29.5%; Pred. No. 9e-30;
Matches 393; Conservative 18; Mismatches 551; Indels 369; Gaps 56;

QY 1 GATCTGGGTGTCTGCC-----AGGAAAAGCAAAATCTG-GAAGTTAATGTTTTCAGTG 54
DB 134 GLLFNGNGNAGAGVGVGGAGGAGLFGIGAGGAGGAGAPGCTGGTGGWLAGGGVG 193
QY 55 ATTTTAAATCTTGTGCGGAGAGGCCCTCTCCCGTATCAGCGCTTCTCTCATT 114
DB 194 G-----MGAGG-GAGG-----AGGNAGLFGNG----- 215
QY 115 CTTGTAATCCGGCTCCCGCTCTTC-----GGCTCAGACAGCCG--GAGGAAGCC 166
DB 216 ----GAGGAGGAG-----GAGGAGNAGWFGHGGAGGVGVGAANGATPGDGAAG-V 266
QY 167 TGTTCGAATTTAAGCGGCTGTGAACGCCCGAGGCGC--GGCGGGGCGGCGGAGCGG 224
DB 267 AGSDDG-AGDGLAGSDGDDGGAGGVGGNGRGRGMLLNGGAGGAGGVGGVGGAGGAAGA 325
QY 225 GGCCATTTTGA--TAAAGAGCGTCCCTTCCAGGCGAGCTCTATAAGTGACCGC----- 277
DB 326 GGAGATGNGPAGTSAAGDGGAG-----GNGAGNGVGGAGGAGSAGLLGYVG 377
QY 278 -CGCGCGAGCGTGCCTGGCGGTGTCAGTCACTGTACGCACTTCTTTTGTCTTCTTCT 336
DB 378 RAGDGGAGGGGLGAPGDDGAGGNGGSLAAGDGGAGG-----HGDDP-----G 422

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1660 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:92261507; PIDN:CAAL16067.1; PID:92791
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv2490c
C:Superfamily: collagen alpha 1(IV) chain


```
337 CTTTGGGACCTCTGAGTCACTCCCA---GATGAAGCGCTGAGCCC---GGTGC 389
423 LGGAGGAGGA---SGGAGAGANGLAAGNDGFPVSGGKGKGGAGHAPVAGHGNG- 477
390 GCGGTGCTA-CGAGCGGTGTCTGCC---TGTCG---GAACGAGTCTGG 434
478 GAGGNGLVGDGAGHGGHGAAGYADMTAIFLGSSGTPGPDGNGHAGGAGG---AGG 535
435 CCAATCCCGCGGCGCGAGG---GAAGCGCGCGGAGCTGAGG---473
536 AHAGDGGAGGAGGAGGAGGAGHGFNAVLVSDGNGGDDGAGGSGGPGGAGGAGDA 595
474 AGCGCTGAGTCTCTGAGCAGACATGAACCACTGCTACTCCGCTGCG---GG 524
596 PAGRAGSOGV-GDGGAGAGGAPONGS---GGRDNVAFKDDGGAGGDDGDPFAGKGG 652
525 AACTGTTACCGGAGTCCGAGAGCACTCAGCTTAGCCAG---GTGGAAT 573
653 AGGAGATEVGTGATGATVHSGGKGKGGAGADATVAGANGKGGAGGNGLVGDGAGGD 712
574 CTTACAGCGGTCATCGACTACATCTCTGACCTCAGGTAGTCTGSCCGAGCCGCC 633
713 GSGAAGANGNVBDGADLTGDPGEGSEANGGGVGG---GAGAGGDDGAGGS 768
634 TGGACCCCTGATGCCCGCCACCTTCCCATCCAGACCGGAGCTCGCTCCGAACTGT 693
769 ALG---SGNGG---RGDAQAGGAGAG---GAGGAGSVSDGG 805
694 CATCTCCACGACAAAGAGCTTTTGCCACTGCTGGCGGTGCTCTGACACCTCCAGA 753
806 ---PGKKGAGAG---GAGAGG---GGKAGSAG---DSAE 836
754 AGCGAGGTCTGCGCCCGTCTGCTGCTGGAGCCCG---GGAACCTCTCTGCGCG 806
837 AVGGAGGCGDGGVGGVGGDGGPGDGGAGGAGAPAGVGHGNGVGG---GDGLGG 890
807 AAGCGG---GAGCGAGGATGGGCCCACTTCCGCTCCCGCTGCTGCTGCTTCTC 856
891 AGNGGDDGSHGSDGGDGGDGGDPGAGGLG---LGDSNGTAAAGVDSADHG 941
857 ACCAATCCCTCTCTGAGACTAAACCTGTGCTCAGAGCGAAG-CACTGTGAATTTGT 915
942 PGSGG---NGNGGNGAQAASVAGGAGGNGDGNAGRVGDGAGGNGGDDAAGA 992
916 AGCTGGAAG---AGCGAGCTAGCTCTGCGCCACCACTGGCGGAC 958
993 NGANGAPGSDALALGPNGNGGGDAGAGAGGAGGAGGAGSVSDGGAGNG---1048
959 GTCACCTCTCCCAACCCCACTTCTAAGTCTTTTTCAGAGC-----GT 1008
1049 -----GAGNGGVGASGAGGANGIDSIGT 1076
1009 GGAGGTGTGGAAGG---AGTGGCTGCTCTCCAACTATGCCAAGCGCGGCGAGAGCTG 1064
1077 GAGGGGGDGGAGVGHGGDGGVGG---AAPSCTVSHGTGGVGGGLGGAGGVG 1130
1065 GTCTCTGCTCTCTTGGAGAAAG-----GTTCTGTGCTCTGATTTATGAAT---C 1114
1131 G-----AGNGGGIGITVGGAGGAGNGDPPGAGRGGLGSDGNGTSAANGVDASKHG 1186
1115 TATAATAGATATAGTTTGTACCTTTTTCAG---GAAAGGTGACTTTCTGTAA 1170
1187 GDDGGVGNAGAKAAAGDGGGG---GDGNAGLFGDGGAGGDA-----DGTAAE 1234
1171 AATGGATGTA 1181
1235 ALGGDGGAGGA 1245
```

RESULT 13
H70846

hypothetical glycine-rich protein Rv3345c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70846
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Baitelli, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; PMID:98295987; PMID:9634230
A:Accession: H70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1538 <COL>
A:Cross-references: GB:AL021841; GB:AL123456; NID:93261517; PIDN:CAA17117.1; PID:g3261:
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv3345c
C:Superfamily: collagen alpha 1(IV) chain

Query Match 10.6% Score 797; DB 2; Length 1538;
Best Local Similarity 28.5%; Pred. No. 1.2e-29;
Matches 410; Conservative 27; Mismatches 553; Indels 450; Gaps 61;

```
QY 6 GGGTGTCTGCCAGGA---AAAAGCAAAATCTGGAAGTTAATGTTTGTG-----50
DB 200 GAGGTGANGGAGGNAWLFAGGSGAGTNGVGGSGGFVYGGGAGGIGGIGGNGGD 259
QY 51 -----AGTGATTTTAAATCTTGTCTGGCGGAGA---GGCCGCGCTCTCCCGGT 97
DB 260 AGLFNGGAGGAGAGLPGAGLNGDSDGNGGTGNGRGGLLVG-----NGGA 311
QY 98 ATCAGCGCTTCTCATCTTTTGAATCCGCGG-----CTCCGCGGTCTTCGCGCTCA 148
DB 312 GAGGCGV-----GDGKGAGDPSFAVNNAGGN---GGHGNP 347
QY 149 GACCAGCGGAGGAGGCTCTTTGCAATTTAAACGG-----184
DB 348 GV-----GGAGGAGLLAGHAGATPTSGNGDGGIGATANSPLAAGGNGNGHG 401
QY 185 ---CCTGTGAACGCCA---GGCCCGCGGGGGCGGGGCC---GAGCGCG 225
DB 402 GLVNGGTGAGGAGHAGSTGATGTALOPTGNGTNGAGHGGNGNGGAQHGGGVGG 461
QY 226 GCATTTTGAATAAAGAGCGCTCTCCAGGCGAGCTCTATAAGTACCGCGCGG---281
DB 462 ---KGGAGSGGAGNGFP-AATLSPGADCGWNGGKGGDGGKAGDGAAGD 513
QY 282 ---GCGAGCGTCCCGCTTGCAGGTCACTGTAGCGGACTCTTT-----TGCTTT 329
DB 514 VTLAVNQAGDGGNGGEGVGGKGGAGVSNPALNGSAGANGTAPTSGGNGNGGAGA 573
QY 330 TCTTTCTCTTTGG-GGCACCTCTGGACTCACTCCCGAGCATGAAGCGCTGAGCCCGGTG 388
DB 574 TPTVAGNGGAGGNGHGGSVNGGA-----GGAGNGVAGTGLALNGN 618
QY 389 CGCGGC---TGCTACGAGCGGTGTGCTGCTGCGAAGC-----CAGTCTGGCATCGC 441
DB 619 GNGGIGGNGGSAAGTGGDGGK-----GNGGAGANGQDFASANGANG---QGG 666
QY 442 CCGGGCGCGAGG---GAAGCGCGCGGAGCTCAGGAG-----CCGCTCAG 483
DB 667 NGNGGGGKGGDAFAFAKAGNGAGGN---GNNVVGAGGAGGKGAIPAMKATGAD 723
QY 484 CTTCCTGAGACATGNAACAC-----TGCTACTCCCGCTGCGGAGC-----527
DB 724 GTAPTSGDGGNGNGGASPTVAGNGGDDGKGGGNGVNGNGGAGNGAGAGQAGTGP 783
QY 528 ---TGGTACCCCGAGTCCCGAGAGGCACTCAGCTTAGCCAGGTGGAATCTTACAGCGG 584
DB 784 TSGDGTSGTDGGAGG---NGGAGG---AGGTLAG---HGGNGKGGGGGGGIG-GAG 832
QY 585 TCATCGACTACATCTCGACCTGC-----AGTACTCTCTGCGCGGAGCCCGCTGGAC 638
```


QY 1003 GAGCTGG-----AGGTGTGAAGAGTGGCTGCTCTCCAACTATGTCACCAAGCGCGGCA 1058
 Db 1144 GGGGCGGDRAGRGDAGLGGSGPGGTGPDWG-----TGTTGGTGGTG 1186
 QY 1059 GAGCTGCTCTCTGCTCTCCCTGAGAAAGTTCTCTGTCCTGATTTATGAACCTATA 1118
 Db 1187 GQANGG-----LTGG-----RGTTGGNGNGNTGTTGGAGGTG-----GTGNGSQPMG 1232
 QY 1119 ATAGAG-----TATATAGTTTGTACCTTTTACAG-----1151
 Db 1233 GNGAGGFGNGFAGVGRGGMGSGGTGTTGAGPFGTGTGTTGGHGGGCGGGSILL 1292
 QY 1152 --GAAGTGACTTCTCTAACAATGCGATG 1179
 Db 1293 GLGGLGGLSPGSATGTAGAGGGGFGG 1322

RESULT 15
 H70987
 hypothetical glycine-rich protein Rv1759c - Mycobacterium tuberculosis (strain H37RV)
 N:Alternate names: wag22 antigen homolog
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: H70987
 R:Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:9829598; PMID:9634230
 A:Accession: H70987
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-914 <COL>
 A:Cross-references: GB:295890; GB:AL123456; NID:G3242245; PIDN:CAB09322.1; PID:G2131027
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv1759c; wag22
 C:Superfamily: elastin

Query Match 10.4%; Score 784.5; DB 2; Length 914;
 Best Local Similarity 29.5%; Pred. No. 3.4e-29;
 Matches 337; Conservative 18; Mismatches 386; Indels 401; Gaps 48;

QY 1 GATCTGGGCTCTCCAG-----GAAAAAGCAATCTCGAAGTTATGTTTGTAGTGA 55
 Db 116 GNGANGAPGTGANGGDAGLWLTGNGAGGSGAKGA---NGGAGPGGAAGLFGNGAG-GA 171
 QY 56 TTTTAAATCTCTGTCGGGAGAGGCCCGCTCTCCCGGTATCAGCGCTTCTCTCATTC 115
 Db 172 GGTATA-----NNGIGAGG-----AGGSAMLFAG-----197
 QY 116 TTTGAATCCGGGCTCCCGCTCTTCGGCGTCAAGCAG-----CCGAGAGAGCCT 167
 Db 198 ---GAGGAGGAATSLVGGIGGTGTGGNAGLAGAGGAGGAGGFSFSTAGGAGAGG---251
 QY 168 GTTTCCAATTTAAGCGGCTCTGAACGCCAGG--GCGCGCGGGCGGCGGAGCGG 225
 Db 252 ---AGLFTTGGVAGAGGQHTGGAGGAGGAGLFG-AGNMGAGGFGDHTLTGTGAGG 307
 QY 226 GCCATTTTGAATAAGAGCGCTGCTTCCAGGCGGCTTATATAAGTCAACCGCGCGCGCA 285
 Db 308 -----DGGGGLFGAGG-----DGGAGGSLTTGGAAGNGNAGTSLGAAAG- 349
 QY 286 GGTGCGGCGCTGTCAGGCTACTAGCGGACTCTTTTGGTTTCTTCTCTTTGGGGC 345
 Db 350 -----GAGTGGAGGTTFGGGKGAGGA-----GNN- 375
 QY 346 ACCTCTGACTCACTCCCGCATGACGAGCGCTGAGCCCGGTGCGCGCTGCT--ACGAG 403
 Db 376 -----AGMLFGSGGGGTGGFPAAGGQG-GVGSAGMLSGSGSG 415

QY 404 GCGGTGTGCTGCTGTTCGGA--ACGCAGTCTGCCATCGCCCGGGCGGCGGAA---G 457
 Db 416 GAGSG-GPAGTAAGAGAGGAGPGLIG--NGNGNGNGSGGTGGVGGAGNAVILNG 472
 QY 458 GCGCCGCGAGCTGAGGAG--CCGTGAGCTTC-----TGACGACAT 498
 Db 473 GEGGIGALAGKSGFGGFGGLLLGADGYNAPESTSPHNLQDILSFINEPEALJGRPLI 532
 QY 499 GAACCACTGTCTACTCCCGCTCGCGGAACCTGGTACCCGAGTCCCGAGAGCACTCAGCT 558
 Db 533 GNGDSGTFG-----TGDDGA--GWLFGNGNG-----GAGAAGTNGSAGGAG 574
 QY 559 TAGCAGGTGAAATCCTACAGCCGCTCATCGACTCATCTCTCGACCTCGAGGTAGTCTCT 618
 Db 575 GAGGILFTGCGAG-----GAGGVGTAGA-----CGAGGAGSAGFL 609
 QY 619 GCGGAGCCAGCCCTCGACCCCTGATGGCCCCCACCCTTCCATCCAGACAGCCGAGCT 678
 Db 610 IGSGTGTVGGAATTTGG-----VGAGG-----NAGLLIAGALG-- 645
 QY 679 CGTCCGGAATCTGTCTATCTTCCAAAGCAAAAGAGCT-TTTCCTGCTGAGTCCGCGGTG 737
 Db 646 ---GCGGGAFTAG-----VTTGGAGGTGGAAGLFGANGA---GGAGGTG 683
 QY 738 TCCTGACACCTCCAGAACG--CAGTGTCTGCGCCCTTCTGCTGCGGACCCCGGGAACC 795
 Db 684 STAGAGG-----AGGAGGLYAHGTGPGNGSGSTGAGTGGAGGPGGLYAGGS--- 734
 QY 796 TCTCTGCGGGAAGCCGAGCGGAGGATGGCCCCCAACTTCGCCCTGCCCACTTGACTT 855
 Db 735 -----GGA---GGHGGVAGGGGVG---NAGSLTL 759
 QY 856 CACCAATCCCTTCTGAGAGACTAAACCTGTGTCTCAGGAGCGAGGACTGTGAATTGT 915
 Db 760 NA-----SGAG-----GSGGSSLSGKAGAGAG--GS 785
 QY 916 AGCCTGAAGAGCCAGAGCTAGCTCTGGCCACCAGCTGGCGACGCTCACCTGTCTCCACC 975
 Db 786 AGLFVSGGAGNGGYSLNG---TGG-----DGGTGGAGQ---817
 QY 976 CCACCCCAAGTTCTAAGTCTTTTACAGCGTGGAGGTGTGGAAGGAGTGGCTGCTCTC 1035
 Db 818 -----ITGLRSFG-GAGGAGASDTGAGGNGGAGGKAG-----850
 QY 1036 CAAACTATGCCAAGCGCGGCGAGAGCTGTCTTCTGCTCTCTTGGAGAAAGTTCTGT 1095
 Db 851 -----LYNGGDDGAGGDDGATSGKAGAGNAVILNG-----GNGGNAGKAGGTAGAG 899
 QY 1096 TG 1097
 Db 900 AG 901

Search completed: May 7, 2004, 15:10:51
 Job time : 53 secs

```

QY 841 CACCGACTGAGATATGTTTACAAATAGTCTG-----TGGGCTGTTTGTGTTATT 895
Db 1070 LGPGGNGGTGNASATGTNGVANAGNGKGGGGGQFAGGNGGAGSVTDGSGAGTAGN 1129
QY 896 AACAAATATTT-----TAGATGGTGAATAAAA 923
Db 1130 GGGNGNATGTIAGQAPAGNGSAGGKGDDGGNIAAGA 1166

RESULT 15
O88281 PRELIMINARY; PRT: 1574 AA.
AC O88281;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MEGF5.
GN MEGF5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9633030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RL like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
DR EMBL; AB011532; BAA32462.1; -.
DR PIR; T13954; T13954.
DR HGSP; P00736; IAP0.
DR GO; GO:000509; F:calcium ion binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000152; ASX_hydroxyl_S.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00008; EGF; 20.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00179; EGF_CA; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00022; EGF_1; 23.
DR PROSITE; PS01186; EGF_2; 23.
DR PROSITE; PS01187; EGF_CA; 5.
KW EGF-like domain.
SQ SEQUENCE 1574 AA; 165445 MW; 2848533D8F77F6E7 CRC64;

Query Match 13.3%; Score 764; DB 11; Length 1574;
Best Local Similarity 28.9%; Pred No. 3.4e-35;
Matches 302; Conservative 31; Mismatches 401; Indels 312; Gaps 74;

QY 1 GGGGCCCATCTCTGTT---TCAGCCAGTC---GCC--AAGATCATGAAGTC--GC--- 46
Db 549 GWTGIIICNETCPDPTGKNCSSPFC--TCONGTCDPVLGACRCPPGVSGAHCEDCPKGF 606
QY 47 -----CAGTGGC---AGCAC-----C-----GCCACCGC----- 67
Db 507 YGKHCKKCHCANRGRCHRLYGACLCDPGLYGRFCHLACPPWAPGCGSEDCLEQSHTR 666
QY 68 -C-----GCCGGGGC-----C-----CAGCTGGCC--GTTGAAGGGCGGCAAGACAGCGAGC 113
Db 667 SCNPKDGSCKAGFQGERCOAECSGFFPGFGRHRTCPQGVACDPVSGECRTQCPPGY 726
* QY 114 GGTGGCGGCGAGGTGTGGCTGTCTGTCTGAGCAGAGCG--TGCCCATCTCGGGCTGCC 171
Db 727 QGEDCCQEC---PVGTGFGVNCSSCS-----CVGAPCHRVG-----ECLCPGKT--- 769
QY 172 GGGGCGC-CGGGCGCGC---CTGCTGCC---TG-----CTGGACGAGCAG-CAGGTAAA 220
Db 770 -GEDCGADCPGRWGLGCOEICPACBHGASCNFETGTCLCLPFGVGRCPDTCAGW--- 825

```

Query Match 13.3%; Score 764; DB 11; Length 1574;

Best Local Similarity 28.9%; Pred No. 3.4e-35;

Matches 302; Conservative 31; Mismatches 401; Indels 312; Gaps 74;

QY 1 GGGGCCCATCTCTGTT---TCAGCCAGTC---GCC--AAGATCATGAAGTC--GC--- 46

Db 549 GWTGIIICNETCPDPTGKNCSSPFC--TCONGTCDPVLGACRCPPGVSGAHCEDCPKGF 606

QY 47 -----CAGTGGC---AGCAC-----C-----GCCACCGC----- 67

Db 507 YGKHCKKCHCANRGRCHRLYGACLCDPGLYGRFCHLACPPWAPGCGSEDCLEQSHTR 666

QY 68 -C-----GCCGGGGC-----C-----CAGCTGGCC--GTTGAAGGGCGGCAAGACAGCGAGC 113

Db 667 SCNPKDGSCKAGFQGERCOAECSGFFPGFGRHRTCPQGVACDPVSGECRTQCPPGY 726

* QY 114 GGTGGCGGCGAGGTGTGGCTGTCTGTCTGAGCAGAGCG--TGCCCATCTCGGGCTGCC 171

Db 727 QGEDCCQEC---PVGTGFGVNCSSCS-----CVGAPCHRVG-----ECLCPGKT--- 769

QY 172 GGGGCGC-CGGGCGCGC---CTGCTGCC---TG-----CTGGACGAGCAG-CAGGTAAA 220

Db 770 -GEDCGADCPGRWGLGCOEICPACBHGASCNFETGTCLCLPFGVGRCPDTCAGW--- 825

Search completed: May 7, 2004, 15:19:15

Job time : 57 secs

```

QY 221 CGTGC-TGCTCTACGAC-ATGAACGGCTGTGTACTCAGCC-----CTCAAGGAG 266
Db 826 YTGCGQIRACANDGHCDPTTGRCSCAPGWTGLSCORACDSGHWGPDCIHPCNCSAGHN 885
QY 267 TGTGTGCC-----C-AC-----CCTGCCCCAGAAAC-----CGCAAGGTGA 300
Db 886 CDAVSGLCLEAGVEGPRCEQSCRQGYVPSCQKCFCEHGAACDHVSGACTCPAGWRGS 945
QY 301 -----GCAAGTGGAGATTCTCCAGCAGCTCATCGACTACATCAGGAGCACTTCAGTTGGAG 356
Db 946 FCEHACPAFPGLDCDSACNCSAGAPCD--AVTGSC-ICPAGRWGPRCAQSCPLTFGLN 1002
QY 357 CTGAACCTCGGAATCCGAAGTTGGGACCCCGGGGGCC-----GA-GGGCTGCC-----GG 405
Db 1003 CS-QICTCFNGASCDSVT-----SQCHCAPGWMGPTCLQACPEGLYGNKQHSCLCRNGG 1056
QY 406 TC-CGGGCTC-----CGCTC-----AGCACCTCAACCG-----CGAGAT 439
Db 1057 RDPILGCTCPEGWTLACNECLPHYAAGQNLNCSLIH-GGICDRLTGHCLCFAGWT 1115
QY 440 CAGC-GCCTTGACGGCC-----GA-----GGCGGCATG-----CGTTCTCTG-CGGA 478
Db 1116 GDKQSCSVSGTFGVHCEHCACRKGASCHHVTGAC-FCPPGWRGPHCEQACPRGWFGEA 1174
QY 479 CGATCGC---ATCTTGTGTGC-TGAAGCGC-----CTCCCCCAGGACCGGC 522
Db 1175 CAQRCLCFPTNASHHVTGSCRPFTGLSCQACQPGTFGKDCHEILCQCPGETWACDPA 1234
QY 523 GGACCCAGCCATCCAGGGGCAAGAGAAATTAGTGC--TCTGTGGGTC---TCCCCCA 577
Db 1235 SGVCTCAAGVHGTGCL---QRCPSG-----RYPGCEHICKLNGGTCDPATGACYCP 1284
QY 578 ACGGCG-CTCGCCGATCTGAGGGGAGAACAGACC--GATCGGGCGGCAC---TGCGC- 629
Db 1285 AGFLGADCSLAC-----PQRFSGSAHVCAACRQGAACDPVSGACICSPKTVRCE 1336
QY 630 --CCTTAACATCCAGCCCTGGGGCTGA-GGCT--GAGGCACTGGCGAGGAGAGCGG 683
Db 1337 HGPCQDRFGKGCCLKCA--CRNGGLCHATNGSCSCLPLGWMGPHCEHACPAGRYGAACLE 1394
QY 684 CTC-----CTCTCTGCAC-----ACCTACTAG-----TCACCAGA-----GACT 717
Db 1395 CFCQNGSCBPTTGACLCGPGFYQACEHSCPSFGHPGQRCVCECQOQAPCDPVSGQL 1454
QY 718 TTAGGGGTGGGATTCACACTCGTGTGTTTCTATTTTGAAGACGACATTTTAAAAA 777
Db 1455 CPAGFPHG-----QFCEKGCESGSGDGLQCNCHTGVPCDPISGLC-----LCPPGR 1502
QY 778 TGGTCACGTTTG--GTGCT-TCTCAG 800
Db 1503 TGAACDLDCRRGRFGPGCALRCDGG 1528

```



```

QY 110 -GAGCGGTGCGG-----GGAGGTGTGCGCTGTCTGTGACGAGAGCGTGCCTAT 160
Db 222 GNGDGGDGAAGAVGISGATAGDGGGGGTGAAGNGGTGAGAGGGIDGVGGGTG----- 277
QY 161 CTGCGGTGCGG-----GCGCGCGGCGCGCTGCTGCTGCTGACACGAGCAGCAG 215
Db 278 -TGNGGNGAIGAGGAGGAGGNSGNG-GIG-----GKGNAGAGGAAGSNG 323
QY 216 GTAAAGTGTGTCTTACGACATGAACGGCTGTCTTACGCTCAAGAGTGTGTGCCC 275
Db 324 GTVGANGTGGDGGNGGAAGAAATASNGAGTG-----SAGNGGTGTGNG----- 367
QY 276 ACCTGCCCCAGAACCGCAAGGTGAGCAAGTGGG----- 310
Db 368 ---GSGGAGGDTGGVGGGKGNGAGDEVGAGGAGGSGPNTSPGNGGCGGGGGAG 424
QY 311 GATTCTCCAGCAGCTCATCATCACTACATCAAGGACCTT-----CAGTTGAGCTGAAGTCTG 365
Db 425 GAACGAGGAGGANGTAGNGGCGGAGGTGGAGAAASATNGSGGAGGTGGDGGSGGAGGTG 484
QY 366 GAATCCCAAGTGTGGACCCCGCGGCGCGAGG--GCTGCCGCTCCGCTCCGCTCAGCAC 424
Db 485 GAGGTGAAGDGGGGGGAGGAGGAGGAGGAGGTGGNGNITGGTAGTAGAAGNGGA- 543
QY 425 CCTCAAGCGGA-GATCAGCGCCCTGACCGCGAGCGGCATGCGTTCTTCGCGGAGATC 483
Db 544 ---AGKGGAGCGGTGGGTG--GCGGAGGAGGAGGTGGDRTVGGGTVPAGSGCGGAG 597
QY 484 GCAT-----CTGTGTCG-----CTCAAGCGCTCCCGCCAGGAGC 519
Db 598 GGGAGGAGGAGDGGSGDGDAGTGGNGGNGNRNSNGTGGAGN-----GGGGAN 648
QY 520 GCGGAGACCCAGCCATCAGGGGCAAGAGAAATTA--CCTGTCTGTGTGGTCTCCCCCA 577
Db 649 GGAGG-----AGSGGCTGGNGAGGAGDAGDAGNGNGTNG----- 686
QY 578 ACGCGCTCGCGGATCTGAGGAGAACAGACCGATCGCGGCGCACTGCGCCCTTAAT 637
Db 687 -----GNGGNGIAGMNGNGAGGTGSG--NNGNGSGNGGNGAGMNGSGTSGD 734
QY 638 GCATCCAGCC---TGG-GGCTGAGGTGAGGCACTGGCG----- 672
Db 735 GAGGNGAAGTGTGTGGDGLTGTTGGTGGGCGDGGNGGNGADNTANMTAAGDGGN 794
QY 673 ---AGGAGGCGCTCTCTCTGACACCTACTAGTACACAGAGACTTTAGGGGT 726
Db 795 GGGGFGGAGAGGG-GLTAGANGTGGG-----GAGGNGAIGGHPITDDPGNGGT 849
QY 727 GGGATT-----CCTCTGCTGTTTCTATTTTGT-----AAAAAGCAGACATTT 770
Db 850 GNGGTGTGTGGAGTGLGGTGGDGGNGGNGTGGEGEVGGAGGTGGAAGNGDGGTGG 909
QY 771 TAAAAAATGGTCACTT-----TGCTGCTTCTCAGATTCTCAGGAAATGCTTT 820
Db 910 TGGDGGAGGTGGTGGTGGDGLDPPRVGSGDGGT-----GSGGAAGNGGNGAGAGN 964
QY 821 GTATTGTATTATCAATATATCAACGACTAGAA--TATTGTTTACATAGTCTCTGTTG- 877
Db 965 GNGGTGGAGGTGTGGNGD-AEPVPPGAGGAGGTGCKGTGGNGSGTGGTGG 1023
QY 878 ---GCTGTGTTTTGTTATTAACAATAATTTAGATGTTGAAAAA 921
Db 1024 GGTGGGNGGTGMNGGKGKGTGSGGAGDGGKAPAGGTGGAGDGG 1070

```

RESULT 13
 Q8VIZ1 PRELIMINARY; PRT; 1384 AA.
 ID Q8VIZ1
 AC Q8VIZ1;
 DT 01-01-2002 (TrEMBLrel. 20, Created)
 DT 01-01-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-01-2003 (TrEMBLrel. 24, Last annotation update)
 DE PE_PGRS family protein.

```

GN MT3612.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA DeCher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007163; AAK47970.1; -.
DR TIGR; MT3612; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000084; PE_region.
DR InterPro; IPR002173; PFKB.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSELL.
DR PROSITE; PS00583; PFKB_KINASES_1; 2.
SQ SEQUENCE 1384 AA; 110839 MW; 124AC8AAADD55BD CRC64;

Query Match 13.3%; Score 765.5; DB 16; Length 1384;
Best Local Similarity 31.4%; Pred. No. 2.5e-35;
Matches 349; Conservative 20; Mismatches 469; Indels 273; Gaps 44;

QY 1 GGGGCCCATCTCTTTTTCAGCCAGTCGCCAAGAAATCATGAAGTCTG-----CCAGTGCA 54
Db 141 GGNG---APGAAGSGGNGSGAGLWNGGAGGAGGSGGAGGNGGNGWLFAGGTGGIG 197
QY 55 GCACCGCCACCGCCCGCCG---GGGCCCCAGCTGCGCGCTGAAGCGCGCA-----AGAC 106
Db 198 GTGAFAMGCTGGNGGNGALLIGGGGLGGAGGAGGTG--GGTGGTGGNGGNGALLIGAGV 256
QY 107 ACAGA-CGCTGCGCGCGAGGTGGTGCCTGT-CTGTCTGAGCAGAGCTGCCCATCTCTG 164
Db 257 GGAGGIGGCTGAGGAAGAGGTGGNG--GAGGLFWNGGPGGAGGGG-GGAAGDAASAGG 314
QY 165 CCGTCCCGG-GGCGCGCGG---CGCGCTGCTGCTGCTGACGAGCAGCAGTAA 220
Db 315 TCGKGGCGGCGGTGGAGGAGPVLFGHGGAG-----MCGGCTGGGAGGAGGTGTTVIA 367
QY 221 CCGTCTGCTTACGACATGAACGGCTGTACTACGCGCTCAAGGAGCTGGTCCCGCC 279
Db 368 AGTGGGTGGAGA---GGAAGAGALTSGGLAGVAGGTGGTGGNGADAATAAVG 424
QY 280 -----TCCCCAGAACCGCAAGGTGAGCAAGGTGAGATTCTCCAGCACGTC 326
Db 425 FGANGDPGAGKGGNGGTGGAATGTVAGDGGTG-GRGCTGGAG-----AGNDAGST 477
QY 327 AT-----CGACTACATCAGGACCTTCAGTTGGAGCTGAACCTCG----- 365
Db 478 GNPFGKGGDGTGGAGAGGAAGTNGGAGTGGDGTGGNGGTGGVNGADNTLN 537
QY 366 ---GAATCCGAAGTTGGACCCCCGGGGGCGCAGG-GCTGCCGTGCGGGTCCGCTCA 420
Db 538 PDTGGAGGPGGAGGAGGAA-----GGPGGTGGTGGNGGNGGNGGNGGNGNG- 588
QY 421 GCACCTCAACCGCGAGATCAGCGCTGACCGCGCAGCGCGCATG-----CGTTCTGTC 475
Db 589 -----GNAGNNTNAPVG--GEGGAGDGGAGGAGGAANGGTAGSGTGGVGD 635
QY 476 GGACGATCGCATTTGTGTGCTGAAGCGCTCCCGCCAGGACCGCGGACCCCGCAT 535
Db 636 GGAGGNGGGKAGTNGSNFVVDGAG-----FSGGAGNGGAGGAGGAGGTGG 686
QY 536 CAGGGGGCAAGAGGATACGTGCTCTCT--GGGTCTCCCGCCAGCGCGCTCGCGGAT 593
Db 687 GNGGSDGG-AGGTGGAG-----GNGIPGTGTETPAGGT-----GAKGGD 724

```

QY	704	AGTCACCAGAGAC-	-----TTTAGGGGTGGGATTCACACTCGTGTGTTTCTA	749
Db	1385	AGTAGAAGGAAKGGAGCGGCTGGTGGGAGGDPGAGGTGGBDFTVGGTVPAGSGG	1444	
QY	750	TTTTTTGAAAGCAGACATTTTAAAAATGGTCA-----CGTTGGTCTTCTCAGATT	803	
Db	1445	QCGNAGGGAGCGGAGGCGDGGGDPGAGTGCNGGNGNRSNGTGGAGGN--GGGGANG	1502	
QY	804	TCCTGAGGAATTG-----CTTTGATTTGTTATATTACAATGATCACCGACTGAGAATA	855	
Db	1503	GAGGAGSGGGTGGNGGAGGADGADGAGNGNGNTGNGGNGNG--TAGMGNGGAGTGGG	1561	
QY	856	TTGTGTTT-----ACAATAGTTCTCTGGGGCTGTTTTTTTTGTATTAAACAATAATTAGA	911	
Db	1562	NGGNGSGGNGGAGMGNSGTGSGDGGAGNGCAAGTGGTGGDGG--LTGTGGTGGSGG	1619	
QY	912	TGGTG	916	
Db	1620	TGGDG	1624	
RESULT 12				
ID	OS3557	PRELIMINARY;	PRT; 1079 AA.	
AC	OS3557;			
DT	01-JUN-1998 (TrEMBLrel. 06, Created)			
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	PGRS-family protein (Fragment).			
GN	RV3512 OR MV023.19.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=H37Rv;			
RX	MEDLINE=96299987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,			
RA	Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
DR	EMBL; AL022022; CAA17749.1; ..			
DR	PIR; B70807; B70807.			
DR	TuberCulat; RV3512; ..			
DR	GO; GO:0004420; F:Hydroxymethylglutaryl-CoA reductase (NADPH) . . . IEA.			
DR	GO; GO:0009058; P:biosynthesis; IEA.			
DR	InterPro; IPR002202; HMG-CoA red.			
DR	PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.			
KW	Complete proteome.			
FT	NON_TER			
ST	1			
SQ	SEQUENCE 1079 AA; 81163 MW; A79718CDBC74B97D CRC64;			
Query Match				
Best Local Similarity 14.2%; Score 819; DB 16; Length 1079;				
Matches 340; Conservative 16; Mismatches 462; Indels 249; Gaps 38;				
QY	2	GGGCCCATCTGTTTCAGCCAGTCGCCAAG-AATCATGAAAGT--CCCCAGTGGCAGCAC	58	
Db	106	GTGSAPGTAGTGGD---GGKGNGGIGAAAGTTGTPVGTGSGGTGGSGAGGTGGDGAAN	162	
QY	59	CGCCACCGCCCGCCGGG-----CCCCAGTGCAGCTGAAGCCGCCAAGACAGC----	109	
Db	163	GGTAGAGGAGNGCKGGDGGAGVTSSTAGNSG-GAGSGGKGGDAGAGGAGATPGANGIA	221	

```

QY 314 T-----CTCCAGCAGCTCATCTACATCATCAGGACCTTCAGTTGGAGCTGAATCGGAATC 370
Db 433 SGAGGTNGSGAGGT-----GGGGMGAGAGAGADNPFTGTTGGDGTGGAAGAGAGGA 488
QY 371 CGAAGTTG-----GGACCCCGGGGGCGAGGGCT-----GCCGGTCCGGCTCC 415
Db 489 AGTGGTGGMGTGAGVGA-----GGGGGAGAGAGADADPQATGTTGTFAGAGGA 543
QY 416 GTCACAGCCCTCAACGGCGAGATCAGCGCCCTGACGGCGGAGGGCGG-CATGGTTCTT- 473
Db 544 G-----GAGGSSAGGTNGSGAGAGTGGGGGAGGAGAGADNPFTGTTGGTGGD 591
QY 474 GCGGACGA-TGGCATCTTGT-GTCGCTGAAGCCCTCCCGGAGGAGCCGCGGA--CCCC 529
Db 592 GTGGAAGAGAGAGAGAGTGGTGMGTGAGV-----GAGGCGGAGGAGAGAGS 644
QY 530 AGCCATCCAGGGGCG-----AAGAGGA-ATTAGTCTCTGTGGTCTCCCCCAACGGCG 583
Db 645 SGAGGTNGSGAGGTGGGGGAGGAGAGADNPFTGIG-----GTGGD----- 685
QY 584 CTCGCCGATCTGAGGAGAGAACAGACCGAT-----CGGCGGCCACTGCGCCCTTAAC 636
Db 686 ---GGTGAAGAGAGAGAGAGTGTGTMGTGAGVGGAGGCGGAGGAGADADPQGA 742
QY 637 TGCATCAGCCTGGGCTGAGGTGAGGCTGAGGCTGAGGCTGAGGAGAGGGCGCTCTCTGAC 596
Db 743 TG-GTGAG-----GAGGAGGAGGSSGAGG--TNGSGAGAGTGGGCGAG-----GAGI 787
QY 597 ACCTACTAGTACACAGAGACTTTAGGGGTGGGATCCACTGCTGTGTCTTATTTTGT 756
Db 788 SFGNSNGGI-----GGTGG-----VGTGGDGNAGTCAAGDPKGTGGTGGSGAGSGG 839
QY 757 AAAAGCAGACATTTTAAATAATGTCACGTTTGGTGTCTCTCAGATTTCTGAGAAATG 816
Db 840 ANFNGGTGGTGTGCTGCKGGMGIADGPGDGNAGVGGKGTNGSGSGTGGTGG 899
QY 817 CTTTGTATTGTATATTACATGATCAGGCTGAGATATTGTTTACATATGTTCTGTG 876
Db 900 PGSGGAGTGGTG-----GRGGAGGCGGAGDGGAAAT-GVGGDGD 940
QY 877 GCGCTGTTTTTTTGTATTAAACAAATAATTAGATGTTGAAAAAAA 924
Db 941 GNGGNGNGGTGVSPPGLGG-----AGTGGLGAGAGGGA 977

RESULT 10
ID O17434 PRELIMINARY; PRT; 988 AA.
AC O17434;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor ampullate silk protein Mispl (fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98200471; PubMed=9541398;
RA Colgin M.A., Lewis R.V.;
RT "Spider minor ampullate silk proteins contain new repetitive sequences
RT and highly conserved non-silk-like 'spacer regions'.";
RL Protein Sci. 7:667-672(1998).
DR EMBL; AF027735; AAC14589.1; --
FT NON_TER
SQ SEQUENCE 988 AA; 79082 MW; 461E03DF3F7085D CRC64;

Query Match 14.4%; Score 829.5; DB 5; Length 988;
Best Local Similarity 32.4%; Pred. No. 5.2e-39;
Matches 307; Conservative 16; Mismatches 363; Indels 261; Gaps 39;

8 ATTCTGTTTCAG-----CCAGTCGCCAAGATCATGAATGTCGCCAGTGGCAGCACCGCC 62
QY

```

```

Db 197 ASSVTTLTISYPAVFPASSSGGYAGAGAVAAAGAGAGGYGRGAGGYGGGGYAG 256
QY 63 ACCGCCCGCGGGCCCCAGCTCGCGCTGAAGCCCGCAAGACAGCAGCGGTGCGG-- 120
Db 257 AGAAAAAAGAGAG-----GAGGYGRG-----AGAGAAAGAGAGAGAGYGGQG 302
QY 121 -----GCCAGGTGTGCGCTCTCTCTGTGAGCAGAGCGTGGCCATCTCGCG 166
Db 303 GYGAGAGAGAAAAAGAGAGAGAGGYGRG-AGAGAGAAAGAGAGAGYV-GGGYGA--GAGAG 358
QY 167 CTCGGGGCGCGCGGGCGCGCTCGCTGC-CTGCTGGAGCAGCAGCAGTAACTGC 225
Db 359 AAAAAAGAGSGGAGGYGRGAGAGAAAGAGAGAGSYYGGGGYV-AGAGAAAAAGAG- 416
QY 226 TGCTCTACGATCAACCGCTGTTACTACGCGCTCAAGAGAGTGGTCCCAACCTGCC 285
Db 417 -----AGAGGYGRGAGAGAGAGAAAAA-----GAGAGGAG-YGGQGGYAGAGAAAA 466
QY 286 AGAACCGCAGGTGAGCAAGTGGAGATTTCCAGCAGCTCTCAGCTACTATCATCAGG 345
Db 467 AGAGAGAGAGGYGRGAGAGAGAAAGAGAG-----AG-GYGGQSGYAGAGAAAAAGAG- 518
QY 346 TTCAGTTGCGACTCAACTCG-GAATCCGAAGTTGGGACCCCGGGGCGCGAGGCTGCC 404
Db 519 -----GGAGGYGRGAGAGAGAAAGAGAGAAAGAG-----GGYGGGGYAG- 560
QY 405 GTCCGGGCTCGCTCAGCACCTCAACGGCGAGATCAGCCCTGAGCGCGAGGCG-GC 463
Db 561 ---AGAGAAAAAGAGAGGA-----GGYGRGAG-AGAGVAAGAGAGAGGYGGGGYAGA 608
QY 464 ATGGTCTCTCGGACGATCGCATCTTGTGTGCTGAAGCGCTCCCGCGAGGACCGCG 523
Db 609 GAGAAAAATGAGAGG-----YGRGAGAGAA-----AGAGATGGAG 648
QY 524 GACCCAGCCATCCAGGGGCGCAAGAGGAATTAAGTCTCTGTGGTCTCCCGCAACGCG 583
Db 649 -----YGGQGGYAGAGAGAAAAAGAGAGAGGYGRG-----AGAGAG- 685
QY 584 CTCGCCGATCTGAGGAGAGAACAGCGATCGCGGCGCACTGCGGCCCTTAATGCA 643
Db 686 -----AAAGAGAAAGAGAGAGG-YGGQGGYAGAGAGAGAAAGAGAGGA--- 730
QY 644 AGCCTGG-GGCTGAGGCTGAGGCACTGGCGA-GGAGAGGGGCGTCTCTCTGACAC 701
Db 731 AGYSRGRAGAGAGAGAGAAAGAGAGAGGYGGGGYAGAGAG----- 772
QY 702 CTAGTCACGAGAGACTTTAGGGGGTGGGATTCCTCGTGTGTTTCTATTTTGA 761
Db 773 -----AAAAAGAG-----SGGAGGYGRGAGAGAGAAAGAG- 811
QY 762 CAGACATTTTAAAAAATGTCACGTTTGTGCTCTCTCAGATTTCTGAGAAATTCCT 821
Db 812 YGGQGGYAGAGAGAAAAAG-----AGAGRGGY-----RGA-----GAGYGGQG- 851
QY 822 TATTGTATATTACATGATCAGGCTGAGATATTGTTTACATAGTTCTGTGGG- 879
Db 852 -----YAGAGAGAGAAAA-----AGAGAGGYG 872
QY 880 -----CTGTTTTTTTGTATTATAACAATAATTTAGATGTTGA 917
Db 873 DKEIACWSRCRYTAVSTTSLSSABASSISSAASSTLVSGYLNTAA 919

RESULT 11
QYTWCO
ID QYTWCO PRELIMINARY; PRT; 1938 AA.
AC QYTWCO;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB PS-PGRS family protein.
GN PE_PGRS55 OR MB3541.

```


Db	1425	SSAAAAAAAAAARABYDRAAGSAAAAAAAAAASGAGSGGGYGMWDGGYGS	1484
Qy	251	CTCAGCCTCAAGAGCTGTGTGCCACCTGCCCCAGAACCGCAAGTCAAGTGA	310
Db	1485	AAAAAAAAAASGAGGAGYGYGSDSAAAAAAAAAASGAGAGGG-YGMWDGGY	1543
Qy	311	GATTCTCCAGCAGCTCAT-----CGACTACATCAGGACCTTCAGTTGAGCTGAA	361
Db	1544	GSDSAAAAAAAAAASGAGRGDGGYGSSSAAAAAAAAAARRAGYDRAAGSAAA	1603
Qy	362	CTCGGATCCGAGTTGG-----GACCCCGGGGGCCG-----	394
Db	1604	AAAAAAAAASGAGSGGGYGMWDGGYSDSAAAAAAAAAASGAGGAGGYGYGSDS	1663
Qy	395	-----AGGGCTCGCGTCCGGCTCCCTCAGCACCTCAAC-----GGC	434
Db	1664	AAAAAAAAAASGAGGAGGGYGMWDG--GYGSDSAAAAAAAAAASGAGRGDGGY	1721
Qy	435	GAGATCA-----GGCCCTGACGCCGAGCGGCATCGCTTCCTCGG	476
Db	1722	GSGSAAAAAAAAAARRAGHRAAGCAAAAAAAAAAASGAGSGGGY-----WGDG	1777
Qy	477	G-----ACGATCGCATCTTGTCGCTGAAGCG-----CCTCCCCAGGACCG	520
Db	1778	GYGSDSAAAAAAAAAASGAGGAGGYGMWDGGYSDSAAAAAAAAAASGAGCTG	1837
Qy	521	GCG-----GACCCAGCATCCAGGGG-----CAAGAGATTACGTG	559
Db	1838	GYGMWDGGYSDSAAAAAAAAAASGAGGAGGGYGMWDGGYSDSAAAAAAAAAAGSG	1897
Qy	560	CTCTGTG--GTCTCCCCA-----ACGCGCTCGCCGATCTGAGGGAG--	602
Db	1898	AGRGDGGYGSAAAAAAAAAARRAGHRAAGSAAAAAAAAAASGAGSGGGY	1957
Qy	603	-----AACAGACCGATCGCGGCCACTGCGCCCTTACTGCATCCAGCC	647
Db	1958	GMWDGGYGSAAAAAAAAAASGAGGAGGGYGMWDGGYSDSAAAAAAAAAAG	2017
Qy	648	TGGGGCTCAGG-----CTGAGCACTGCGAGGAGAG--GCGCTCCTCTCTGCA	695
Db	2018	SGAGGAGGYGMWDGGYSDSAAAAAAAAAASGAGRGDGGYGSSSAAAAAAAAA	2077
Qy	696	CA-----CCTACTAGTCACGAGACTTTAGGGGTGG-----ATTCC	734
Db	2078	AARRAGHRAAGSAAAAAAAAAASGAG-----RSGGGYGMWDGGYSDSAAAAA	2133
Qy	735	ACTCGTGTGT-----TCTATTTTGAAGAAGCAGACA---TTTTAAAAAT	778
Db	2134	AAAGSGAGGYGMWDGGYSDSAAAAAAAAAASGAGSGGGYGYGSDSAAAAA	2193
Qy	779	GCTACGTTTGTG-----CTTCTCAGATTCTGAGAAATTCGTTTGATG	826
Db	2194	AAAAAAGSGAGGYGMWDGGYSDSAAAAAAAAAASGAGSGGGY-----YGS	2250
Qy	827	TATATTCAATGATCACCGACTGAGAAATTGTTTCAATAGTTCGTGGGCTGTTT	886
Db	2251	AAAAAAAAAASGAGGYGMWDGGYGYGSDSAAAAAAAAAASGAGGAGGYG	2310
Qy	887	TTTG-TTATTAAACAAATAATTAGATCGTGAAAAA	926
Db	2311	GDSGYGSAAAAAAAAAARRAGHGRSSGAAAAAAAAA	2351

RESULT 6

Q8VIZO

Q8VIZO PRELIMINARY; PRT; 1715 AA.

AC Q8VIZO;

DT 01-MAR-2002 (TReMBLrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE PE FGRS family protein.

GN MT3612.1.

OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
CX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeSoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007163; AAX47971.1; -;
DR TIGR; MT3612; -;
DR InterPro; IPR000084; PE region.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE region; 1.
DR PROSITE; PS00583; PFkB_KINASES_1; 4.
SQ SEQUENCE 1715 AA; 133832 MW; E857B2A774C771B0 CRC64;

Query Match 15.0%; Score 864; DB 16; Length 1715;
Best Local Similarity 34.9%; Pred. No. 9.7e-41;
Matches 335; Conservative 14; Mismatches 436; Indels 176; Gaps 40;

QY 1 GGGGCCCATCTCTTTTCAGCCAGTCGCCACAGATCATGAAGTCGCCAGTGGCGAGCACCG 60
DB 193 GAGGWLFGVGAGGVSAG--GGTGGAGGPGGLIWGGGGAGGVGGAGGTGGAGGERA--E 248
QY 61 CCACCGCGCGCGCGGCCCGCCAGCTGCOCGCTGAAGCGCGGCAAGACAGCG--AGCGGTGC 118
DB 249 LFPAGAGGAGTDDG----PGATG-GTGGHGGVGGDGLAPGAGGAGGCGGAGGAGS 303
QY 119 GGGC--GAGGTGCT----GGCTGTC---TGTCTGACGAGCGTGGCCATCTC-GGCT 168
DB 304 DGGALGTGGTGTGTGGAGGAGGALLLGAGCGGGLGGAGCGGTGGAGDGVGGVGGT 363
QY 169 GCCGG-GCGCGCGCGCGCGCG--CCTGCTGCCCTGTCTGGACGACGACGAGTAACCGTGC 225
DB 364 GKGGVGVGAGLGAGGAGAGOLFSAGAGAGAVGVGGTGGCGGAGGAGAGADAF-STGL 422
QY 226 TGCTCTACGACATG-----AACGCTGTACTCAGCCTCAAGAGAGCTGGTCCCAACCT 280
DB 423 TGGTGFAGGAGGVGGGAGNATGAGTNG-----SGAGGTGGCGGAGGAGGMS 468
QY 281 GCCCGAGAACCGAAGGTGACCAAGGTGGAGATTCTCCAGCACGCTCATCGACTACATCAG 340
DB 469 GADNASGIGADG-GAGGTGGNAGAGAGG-----AAGTGGTGGVVGAGAGAGI 515
QY 341 GGACCTTCAGTGTGGAGCTGGAACCTCGGAATCCGAATGGGACCCCGGGGGCCGAGGCT 400
DB 516 GG-----TGGCGGAGGAGSAGTDATATGATGGTG-----FSGAGGAGGAGGAGN-T 559
QY 401 GCCGCTCCGGCTCCGCTCAGCACCCCTCAAGCGGAGATCAGCGCCCTGACGGCGGAGGC 460
DB 560 GVGGTNGSGGGGTGGAGGAGA-----GGVGADNPTGIGTGTGTGGKAGGAGAGA 611
QY 461 GGCATGCCCTTCCTGGGACGATCGCATCTGTGTGCTGAAGCGCCTCCCCAGGAGACG 520
DB 612 GADATATGATGGTGFAGGAGGAGGGGSSGAGGTNGSGAGGTG-----GGGGA--G 661
QY 521 GCGGACCCCGACCATCCAGGGGGC-----AAGAGGAATTACGTGTCTGTGGGTCTCCCC 575
DB 662 GAGG-----AGADNPTGIGGAGGTGGTGGAGAGGAG--GATGTG----GTGG----- 703
QY 576 CAAACGCGCTCCCGGATCTGAGGAGAACAGACCGATCGCGCGCCACTGCGCCCTTAA 635
DB 704 -AVGVGNAGTGGTGGTGGVGGAGGAGGAGAAAAGSATGAGAGPAGGAGGEGG----- 753
QY 636 CTGCATCCAGCCTCGGGCTGAGGCTGAGGCACTGGCGAGGA-----GAG--GCGGCTCTC 689

```

DE PE PGRS family protein.
GN MT3615.3.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterinase; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE007164; AAK47974.1;
DR TIGR: MT3615;
DR InterPro: IPR002952; Eggshell.
DR InterPro: IPR000084; PE region.
DR InterPro: IPR002173; pFKB.
DR Pfam: PF00934; PE; 1.
DR PRINTS: PR01228; EGGSHLL.
DR ProDom: PD001223; PE region; 1.
DR PROSITE: PS00583; pFKB_KINASES_1; 3.
DR NCBI_TaxID=1773;
SQ SEQUENCE 1217 AA; 94463 MW; 2F5BDCA5A30A056E CRC64;

Query Match 15.4%; Score 889.5; DB 16; Length 1217;
Best Local Similarity 34.7%; Pred. No. 2.8e-42;
Matches 350; Conservative 20; Mismatches 471; Indels 167; Gaps 41;

QY 1 GGGGCCCATCTGTTTTCAGCAGTCGCGCAAGATCATGAAGTCGCGAGTGGCAGCACCG 60
DB 180 GAGGWLFGVGGAGVGAG--GGTGGAGGPGGLIWGGGGAGVGAGGTGGAGGRR--E 235
QY 61 CCACCGCGCGCGCGCGCGCGCGCTGCGCTGAAGCGCCGCAAGACAGCG--AGCGGTGC 118
DB 236 LLFGAGGAGGAGTDDG---PGATG-GTGCHGVGGVGGGLAPGAGGAGGCGGAGGAGS 290
QY 119 GGGC--GAGTGGT-----GGCTGTC-----TCTCTGACAGAGGTGCGCATCTCGG 166
DB 291 DGGALGGTGTGTGTGGAGGAGGAGGALLLGGAGGQGLGGAGGQGGMGAGGAGADNFTIG 350
QY 167 CTGCGCGGCGCGCGCGCGCGCGCTG---CCTGCGCTCTGAGCAGCAGCAGTAA--ACG 222
DB 351 CTG--GDGTTGSSAGG--GAGGAAGQLFSAGAGAGVAGGAGGCGGCGGAGGAGADQ 407
QY 223 TGCTGCTCTACGATGACG--CCTGTTACTACGCTCAAGAGCT-----GGTGCC 274
DB 408 PGATGGTGFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 465
QY 275 CACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 332
DB 466 GADNFTGIGTGGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 523
QY 333 TACATCAGGAGCCTTACATTGAGCTGACCTGCGAATCCGAGTTGGACCCCGCGGGGG 392
DB 524 -----AGGAGADADPQATGTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 573
QY 393 CGAGSGCTGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 452
DB 574 GGGAGSGAGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 625
QY 453 GCGGAGGCGGCA----TG--CGTCTCTCGGA-----CGATCGCATTTGTGTGCTGAA 501
DB 626 GAAGAGGAGGAAGTGTGTGTTTGNAGVGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 685
QY 502 GCGCTCTCCCGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 555
DB 686 G-----GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 732
QY 556 CGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 614

```

```

DB 733 TGIG-----GTGGD-----GTGGAAGAGGAGGAGGAGTGTGTGTGTGTGT 770
QY 615 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668
DB 771 NAGVGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 829
QY 669 GGC--GAGGAGGAGGCGCTCTCTCTGACACCTACTAGTACACAGAGACTTTAG--GGGG 725
DB 830 GGAGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 883
QY 726 TGGGATTCCTCCTGCTGTGTGTCTTATTTTGAAGAGCAGACATTTTAAAAATGTCACG 785
DB 884 TGGTGGTGGGAG--GSGGANFNFGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 942
QY 786 TTTGGTCTCTCTCAGATTTCTGAGGAATTCCTTTGTTTATTTGTA-----TATTA 833
DB 943 -GKGTGGAGDSDSAGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1000
QY 834 CAATGATCACCGACTGAGAA--TATTGTTTACAAATAGTCTGTGGGGCTGTT----- 884
DB 1001 GSGGGGPGFAGGPGGKGGAGGAGTGTGNSGAGGAG---GGGAGGAGGAGGAGGAG 1056
QY 885 -TTTTGTTTATTAACAAATAA-----TTTATGATGTTGTAATAAAAA 924
DB 1057 CTGCTGGVGTGGTGGGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1104

RESULT 5
Q964F4 PRELIMINARY; PRT; 2655 AA.
AC Q964F4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fibroin.
OS Antheraea yamamai (Japanese oak silkworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Saturniidae; Saturniinae; Saturniini; Antheraea.
OX NCBI_TaxID=7121;
RN [1]
RN SEQUENCE FROM N.A.
RC TRANSPOSOME=mariner-like element;
RA Lee J.-S., Hwang J.-S., Kang S.-W., Suh D.-S., Jin B.-R., Kim Y.-S.,
RA Lee K.-S., Goo T.-W., Yun B.-Y., Kim K.Y.;
RT "Complete nucleotide sequence of fibroin gene cloned from Antheraea
RT yamamai."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF325500; AAK83145.1;
DR SEQUENCE 2655 AA; 218672 MW; 89BB810EF9CC4CA6 CRC64;
SQ

Query Match 15.2%; Score 879; DB 5; Length 2655;
Best Local Similarity 26.7%; Pred. No. 2e-41;
Matches 315; Conservative 57; Mismatches 531; Indels 278; Gaps 33;

QY 1 GGGGCCCATCTGTTTTCAGCAGTCG-----CCAGCAATCATGAAAGT 43
DB 1194 GYGSDSAAAAAASAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1253
QY 44 CCGCAGTGGCAGCACCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103
DB 1254 RGDGGYGSAAAAAASAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1305
QY 104 GACAGGAGCGGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 156
DB 1306 AAAAAAGSAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1364
QY 157 ----CCATCTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 209
DB 1365 AAAAAAAGSAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1424
QY 210 CAGCAGGTTAACTGCTGTC-----TCTAGCATGAACTGCTGT-----TA 250

```


QY 61 CCACCGCGCGCGG--GCCCGAGCTGCGCGCTGAAG--GCCGCGAGACAGCGCGT 116
 Db 1033 STGATGATGATGATGATGATGATGATG--STGATGATGATGATGATGATGATGATG 1091
 QY 117 GCGGCGCA-GGTGGG--CGCTGTCTGTCTGACAGACGCGTGGCACTCG--CGCTCCCGG 173
 Db 1092 GATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1150
 QY 174 -GGCGCGCG--GGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
 Db 1151 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1210
 QY 232 AGCATGAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 291
 Db 1211 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1267
 QY 292 GCAAGTGAAGAGTGGAGATCTCCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 351
 Db 1268 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1327
 QY 352 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 402
 Db 1328 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1387
 QY 403 CGGTCGGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 459
 Db 1388 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1443
 QY 460 CGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 519
 Db 1444 ATGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1502
 QY 520 GGCG--GACCCAGCGATCCAGGGGCG--AAGAGGAATTACGCTG--CTGTGGGCTCCCG 575
 Db 1503 GATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1558
 QY 576 CAACGCGCTCGCGATCTGAGGAGACAGACCGATCGGCGGCGACCTGCGGCGCTTAA 635
 Db 1559 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1614
 QY 636 CTGATCCAGCTCGGCGCTGAGGCTGA--GGCACTGGCGA--GGAGAGGCGGCTCTCTG 693
 Db 1615 ATG--GTGATG--GTGATG--GTGATG--GTGATG--GTGATG--GTGATG--GTGATG 1672
 QY 694 CACACTACTAGTAC--CAGAGACTTTAGGGGGTGG--GATTCACCTGCTG--TGCTTCTAT 750
 Db 1673 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1732
 QY 751 TTTTGAAGCAGACATTTTAAATAATGGTCACTT--TGCTGCTTCTCAGATTTCTGA 808
 Db 1733 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1792
 QY 809 GGAATTGCT--TTCTATTGATATTACATGATCACCAGTGAATATTGTT--T 861
 Db 1793 TGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATGATG 1847
 QY 862 TACAATAGTCTG--TGGGGCTGTT--TTTTTTTATTAAACAAATTAATTTAGATGCAAA 919
 Db 1848 GSTGATGATGATGATGATGATGATGATGATG--TGATGATGATGATGATGATGATG 1905
 QY 920 AAAAAA 926
 Db 1906 STGATGA 1912

RESULT 2
 Q98MG7 PRELIMINARY; PRT: 3145 AA.
 AC Q98MG7
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical glycine-rich protein mlr0587.

GN MLR0587.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 CX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimizu S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP002995; BAB48146.1; --
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 38
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 3145 AA; 271616 MW; F10C4A98282766FB CRC64;
 Query Match 17.8%; Score 1028.5; DB 16; Length 3145;
 Best Local Similarity 37.2%; Pred. No. 1.1e-49;
 Matches 369; Conservative 18; Mismatches 499; Indels 107; Gaps 41;
 QY 1 GGGCCCATCTCTTTTCAGCCATCCGACGATGATGATGATGATGATGATGATGATG 55
 Db 1351 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1309
 QY 56 CACCGCACCGCGCGCGG--GCCCGAGCTG--CGCTGCTG--GCCGCGAGACAGC 109
 Db 1310 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1369
 QY 110 GAGCGGTGCGGCG--AGGTGGTGC--GCTGTC--TGCTCAGCAGACGCGTCCCTCTCG 165
 Db 1370 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1429
 QY 166 -GTCGCGG--GGCGCGG--GGCGCGG--GGCGCGG--GGCGCGG--GGCGCGG--GGCGCGG 219
 Db 1430 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1489
 QY 220 ACCTGCTCTACGACATGACGCGCTGTTACTCAGCG--GTGCTGATGATGATGATGATGATG 268
 Db 1490 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1547
 QY 269 GGTGCGCACCTGCGCGCGAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 328
 Db 1548 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1607
 QY 329 CGACTACATCAGGAGACCTTCAGTTGAGCTGAACCTG--GAATCCGAAAGTTGGGACCC 386
 Db 1608 GSTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1660
 QY 387 GGGGCGCGAG--GCTGCGCGTCCGGCTCCGCTCAGCACCTCA--ACGGCGAGATCAG 442
 Db 1661 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1720
 QY 443 CGCCCTCAGC--GGCAGGCGGATCGCTTCTCCGCGAGCAGCAGCAGCAGCAGCAGC 497
 Db 1721 TG--ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1777
 QY 498 TGAAGCGCTCCCGCGGACCGCGGACCCAGCCATCCAGGGGC--AAGAGGATATAC 556
 Db 1778 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1827
 QY 557 GTGCT-CTGTGGGTCTCCCGAGCGCTCGCGGATCTGAGGGAGAGACAGACCGATC 615
 Db 1828 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1882
 QY 616 GCGCGCACTGC--GCCCTTAACATGATGATGATGATGATGATGATGATGATGATGATG 670

Db 261 CPGNCK--NGGACVGVATYNCRCPEVTGQYCTEDVDEQOLMENACQAGTCHNTHGG 318
Qy 217 TAAACGTGCTCT--ACGACATGAACGGCTGTACTC-----AC-- 255
Db 319 YNCVUNGTGDCSENIDDCASACFOGATCHDRVASFYCECPHGRGTGLLCHKACIS 378
Qy 256 -----GC-----CTCAAG--GAGCTGGTGC-----CACCTG-CCCAGAAC 291
Db 379 NPCNEGSCDTPNVNGKRICTCPSTGTPACQDVEDCDLGNRCEHAGKCLNTLGSFEC 438
Qy 292 GCAAGGTGAGCA-----AGGTGAGATTCTCCAG-----CACGT 325
Db 439 QCLQYTGCEIDVNECISNFCQNDATCLDQIGEFQICMPGHEGYCEINTDECASSP 498
Qy 326 CATCGAC-----TACAT--CAGGACCT-----TCAGTTGG 354
Db 499 CLHNGHCKMKIHEFOCCPKFNGHLCOYDVEDACASTPCXGAKCLDGPNTYTCVCTEGY 558
Qy 355 AG-----CTGAACTCGAA--TCCGAAGTTG-----GGAC 382
Db 559 TGTCHCEVDIDCDPDPCHYGSCKDGVATFTLCQPGYTGHHCEINNECHSQPCRHGTC 618
Qy 383 -----CCCCGG-----GGGC-----CGAGGGCTG--C--- 402
Db 619 QDRDNYLCLLKGTTGPNCEINLDDCASNPDCSGTCLDKIDGYECACEPGYTGSMCNVN 678
Qy 403 ---CGGTCCGGCTC-----CGC-----TCAGCACCTCAAC--GGCGAGATCAG 442
Db 679 IDECASPCHNGTCEGDIAGTCTCPGEGHDPFTCLSEVNECNSNPCIHGACRDGLN--G 736
Qy 443 CGCCCTGACGG-----CCGAGGGCGGATCGTTCTCTGGG----- 477
Db 737 YKDCAPGWSGTNCDINNCEBSPCNVNGTCKDMTSGYVCTCREGFSGPNQTNINECA 796
Qy 478 -----ACGATCG-----ATCTTGTGTGCTGAAGGCC-----TCCCC 511
Db 797 SNFCLNQGTICDDVAGYKNCPLPYTGATCEVVLAPCATSPCKNSGVCKESEDYESFCV 856
Qy 512 CAGG-----GACCGCGGA--CCCCAG-----CCATCC 537
Db 857 CPTGWOQTCEVDINECVKSPRHGASCONTNGSYRCLCOAGYTGORNCESDIDDCRPNC 916
Qy 538 AGGGG-----GCAAGAGA-----ATTAGTGTCTGT----- 565
Db 917 HNGSCTDGINAFCDCLPFQGAPEEDINECASNPQNGANCTDCVDSYTCCTCPVGFN 976
Qy 566 -----GGGTC-----TCCCCAACGGC-----C-TCCCGCGATC 594
Db 977 GIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPFTGTSYCOYDVNECDSPCLHGTC 1036
Qy 595 TGAGGGAGAACAGA-----CCGATCGGGCG-----CAC-----TGCGCC 631
Db 1037 QDSYGYKCTCPGYTGLNQNLRKDSAPCKNGRCWQNTQYHCEBSGWTGVNCDV 1096
Qy 632 TTAACGTGCA-----TCAGCGCTGGGGTGA-----GCTGA-----G 663
Db 1097 LSVSCREAAQRGIDVTLCCQHGGLCVDDEGKHCHCOAGYTGSGYCEDEDECSNFCQN 1156
Qy 664 GCACT--GG-----CGAGGAGA-----GGGC-----GCTC----- 686
Db 1157 GATCDYLGFSCKVAGYHGSNCSEINECLSQPCQNGTICDILTNSYKSCSPRGTVGV 1216
Qy 687 -CTCTCTGC-----ACACTACTAGTACAGACATTAG--GGGTGGGATTCCACTC 738
Db 1217 HCEINVDCHPPLDPRASPKFNNGTCDVQVGGYTCCTPPGFVGERCEGDVNECLSNPC 1276
Qy 739 GTGTGTTCTATTTTGAAGACAG--ACATTTTAAAA--AATGTCACGTTTGGTGCTT 795
Db 1277 DP-RGQNCQVQVNDFHCECRAGHTGRRCESVINGCRGPKCKNGSVCAVASNTARGFICR 1335
Qy 796 CTCAGATTCTGAGAAATGCTTGTATTATATTAC--AATGATC-----ACCGA-- 846

Db 1336 CPAGFEGATCENDARTCGSLRCLNGGTCTISGRSPTCLCLGSPGTGPECQPPASSPCVGSN 1395
Qy 847 -CTGAGAATATGTTTACAATA---GTTC-----TGTGGG 879
Db 1396 PCYNGTCEPTSENFYRCLCPAKFNGLLCHILDYSFTGGAG 1437

Search completed: May 7, 2004, 15:18:20
Job time : 32 secs

[9] POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
among mammalian Notch family members."
Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
[10]
RX INTERACTION WITH DTX1 AND DTX2.
RX MEDLINE=21123790; PubMed=11226752;
RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
RA Okano H., Matsuno K.;
RT "Murine homologs of *deltex* define a novel gene family involved in
vertebrate Notch signaling and neurogenesis."
Int. J. Dev. Neurosci. 19:21-35(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
Upon ligand activation through the released cell-cell intracellular
domain (NICD) it forms a transcriptional activator complex with
RBP-J kappa and activates genes of the enhancer of split locus.
Affects the implementation of differentiation, proliferation and
apoptotic programs (By similarity). May play an essential role in
postimplantation development. Probably in some aspect of cell
specification and/or differentiation. May be involved in mesoderm
development, somitogenesis and neurogenesis. Involved in the
maturation of both CD4+ and CD8+ cells in the thymus.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
terminal fragment N(EC) which are probably linked by disulfide
bonds. Interacts with DTX1 and DTX2.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
proteolytical processing NICD is translocated to the nucleus.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=Q01705-1; Sequence=Displayed;
Name=2;
IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
thymus. Expressed at lower levels in the spleen, bone-marrow,
spinal cord, eyes, mammary gland, liver, intestine, skeletal
muscle, kidney and heart.
CC -!- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
endothelial cells, while much lower levels are seen in the
neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
ectoderm, eye and developing whisker follicles.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
which is proteolytically cleaved by a furin-like convertase in the
trans-Golgi network before it reaches the plasma membrane to yield
an active, ligand-accessible form. Cleavage results in a C-
terminal fragment N(TM) and a N-terminal fragment N(EC). Following
ligand binding, it is cleaved by TNF-alpha converting enzyme called
(TACE) to yield a membrane-associated intermediate fragment called
notch extracellular truncation (NEXT). This fragment is then
cleaved by presenilin dependent gamma-secretase to release a
notch-derived peptide containing the intracellular domain (NICD)
from the membrane.
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 36 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC
DR EMBL; Z11886; CAA77941.1; -
DR EMBL; L02613; AAK14898.1; -
DR EMBL; X68278; CAA48339.1; -
DR EMBL; AJ238029; CAB40733.1; -
DR EMBL; X82562; CAA57909.1; -
DR PIR; A46019; A46019.
DR PIR; B49175; B49175.
DR HSSP; P00740; IEDM.
DR MGD; MGI:97363; Notchl.
DR GO; GO:0005887; C: integral to plasma membrane; IC.
DR GO; GO:0005515; F: protein binding; IPI.
DR GO; GO:0030154; P: cell differentiation; IMP.
DR GO; GO:0007386; P: compartment specification; IMP.
DR GO; GO:0007219; P: N signaling pathway; IC.
DR GO; GO:0045944; P: positive regulation of transcription from P...; IDA.
DR InterPro; IPR001152; ASX hydroxyl_S.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch_dom.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 7.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 3.
DR Pfam; PF0002279; Notch; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFBLAMIN.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS00026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Receptor; transcription regulation; Activator; Differentiation;
KW Developmental protein; repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW Alternative splicing.
FT CHAIN 1 18
FT SIGNAL 19 2531
FT CHAIN 1711 2531
FT CHAIN 1744 2531
FT CHAIN 19 1725
FT DOMAIN 19 1725
Query Match 9.6%; Score 553; DB 1; Length 2531;
Best Local Similarity 22.0%; Pred. No. 5.3e-21;
Matches 313; Conservative 38; Mismatches 493; Indels 578; Gaps 78;
QY 28 CAGATCATGAAGTCGCGAGGAG-----CACCGCC-----62
DB 24 CSQPSGTCLNG---GRCEVAGTEACVAGSFVGRQCPNCLSTRCKNAGTCYVDHG 80
QY 63 -----ACCGC-----CGCGCCGCGCGCGCTT-GAAGCC--98
DB 81 GIVDYACSCPLGFLCLTLPDKLANPCRNNGTCDLLTTEYKRCSPGWSGKSCQ 140
QY 99 -GCACAGACAG-----CGAGCGTGC-----GGGCGAGG--TG 128
DB 141 ADPCANPCNAGCGCLPFSSSYTCRPPGHGTCQDVNECSQNFGLCRHGCHNEIG 200
QY 129 GTGCGTGTCTG-----TCTGAGCA-GAGC-GTGGCAATCTCGCTG--C-----170
DB 201 SYRCACATHTGHCFLPYVPCSPQNGATCRPTGDTTTCACLPFGAGQCNCEVDD 250
QY 171 CGGGGCGCGCGCGCGCGCTTGC-----TGCCCT-----GCTGGAC-GAG-CAGCAGG 216

Db 438 CQCQGYTGPRCEIDVNEICSNPCQNDATCLDQIGFEQCLMPEYGVYCBINTDECLASS 497
QY 325 TCATCGACT-----ACAT--CAGGCACT-----TCAGTTG 353
Db 498 PCLNRCVCKINEFLCQCPKFSGLHCQYDVDECASTPCCKNGAKCLDGNPTTCVCTEG 557
QY 354 GAG-----CTGAAGTGG-----GAA--TCCGAAGTGG-----GGA 381
Db 558 YTGTHCEVDIDECDFPCHGLCKDGVATFTCLCQPGYTGHHCEHNECHSQPCRHGT 617
QY 382 C-----CCCCCG-----GGGC-----CGAGGGCTG--C-- 402
Db 618 CQDRNYYLCLCLKLTGTPNCEINLDBDCASNPCDSGTCLDKIDGYECACEPGYTGSMCNV 677
QY 403 ---CGGTCCGGGCTC-----CGC-----TGAGCACCTCAAC--GGCGAGATCA 441
Db 678 NIDECAGSPCHNGTCEGDIAGFTCRPEGYHPTCLSEVNECNSNFCIHGACRDLN-- 735
QY 442 GGGCCCTGACCG-----CCGAGGGCGGATCGGTTCTCTGCGG----- 477
Db 736 GYKCDCAPEWSGTNCDINNNECESNPNVNGTCKDMTSGYVCTCREGFGPNCOINIBC 795
QY 478 ---ACGATGCG-----ATCTTGTCGCTGAAGCGCC-----TCCC 510
Db 796 ASNPCLNQTCLDDVAGYKNCPLPYTGATCEVVLAFCATSPCKNSGVCKESDYEFSC 855
QY 511 CCAGG-----GACCGGCGGA--CCCCAGCCATCCAG----- 539
Db 856 VCPGTGQGTCEIDINECVKSPCRHGASQNTGYSYCLCQAGYTGRCNESDIDCRPNP 915
QY 540 ---GGGGCAAGAGGA-----ATTACGTGCTCTG-----TG- 566
Db 916 CHNGSCTDGVNAFCDCLPGFQAFCEEDINECATNPQNGANCCTCDVDSYTCCTPTGF 975
QY 567 -----GGTC-----TCCCCCAACGCC-----C-TGCGCGAT 593
Db 976 NGIHCENNTPTDESSCFNGTGDVGINSTFCLCPGFTGYSYQYDWNEDSRPCLHGT 1035
QY 594 CTGAGCGAGAACAGA-----CCGATCGCGCG-----CAC-----TGCGCC 630
Db 1036 QDSYGTGKTCPOGYTGLNQLVRNCDGAPCKNGKQWQNTQVHCERSGWTGFCND 1095
QY 631 CTTAACTGCA-----TCCAGCTGGGGCTG-----AGCTGA----- 662
Db 1096 VLSVCEVAAXRGIDVTLQCHGLCVDBEDRHYCHQAGYTGSCYCEBDEVBCSPNQC 1155
QY 663 GGCACCT--GG-----CGAGGAGAG-----GCG-----GCTCTCTCT--GCACACCTA- 701
Db 1156 NGATCTDYLGGFSCKVAGYHGSNCBEBECLSQPCQNGGTCTDILNTYKCSGPRGTG 1215
QY 702 --CTAGTCAC-----CAGACACTTAG-----GGTGGGATTCACCT 737
Db 1216 VHECINVDCHPDPASRSPKCFNNGTCDVQGVYCTCPGFGVGERCEGVNECLSNP 1275
QY 738 CGTGTGTTCTATTTTGAAGAGCAG-ACATTTTAAAG--AATGTCACGTTTGTGCT 794
Db 1276 CDP-RTGQVQVNDPHECRAGHTGRCEVINGCRKPKRNGGVCAVASNTARFGIC 1334
QY 795 TCTCAGATTTCTGAGGAATTCCTTTGTTATGTATATAC--AATGATC-----ACCA 846
Db 1335 RCPARFEGATENDARTCGSURCLNGGTCTISGPRSPCLCLGSGFTGPECQFPASSPCVGS 1394
QY 847 --CTGAGATATGTTTACATAA---GTTT-----TGTGG 877
Db 1395 NPYCNGTCEPTSESFPYRCLCPAKENGLLCHILDYSFTGAAG 1437

RESULT 15

NTC1 MOUSE

ID NTC1 MOUSE

AC Q01705; Q06007; Q61905; Q99JJC; Q9QW58; Q9R0X7;

DT 01-NOV-1995 (Rel. 32, Created)

PRT; 2531 AA.

DT 01-FEB-1995 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Metch A) (MT14) [p300].
GN NOTCH1 OR MOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX TISSUE=Embryo;
MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A., Copeland N.G., Gridley T.;
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of Drosophila Notch.";
RL Genomics 15:259-264(1993).
RN [2]
RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RX STRAIN=CD-1; TISSUE=Embryo;
MEDLINE=93050801; PubMed=1426644;
RA Resame A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
RT "Expression analysis of a Notch homologue in the mouse embryo.";
RL Dev. Biol. 154:377-387(1992).
RN [3]
RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RX TISSUE=Embryo;
MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M., Greenspan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development.";
RL Development 115:737-744(1992).
RN [4]
RP SEQUENCE OF 1161-1547 FROM N.A.
RX STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [5]
RP SEQUENCE OF 1659-1673 FROM N.A.
RX MEDLINE=93364499; PubMed=10437788;
RA Lee J.S., Ishimoto A., Yanagawa S.I.;
RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
RL FEBS Lett. 455:276-280(1999).
RN [6]
RP SEQUENCE OF 1950-2201 FROM N.A.
RX MEDLINE=98029436; PubMed=9384671;
RA Messierie M., Folio M., Nehls M., Eggert H., Boehm T.;
RT "Dynamic changes in gene expression during in vitro differentiation of mouse embryonic stem cells.";
RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN [7]
RP SEQUENCE OF 1655-1659. CLEAVAGE BY FURIN-LIKE CONVERTASE, AND MUTAGENESIS OF 1651-ARG-ARG-1654.
RX MEDLINE=98318619; PubMed=9653148;
RA Logeat F., Bessia C., Brou C., Lebail O., Jarriault S., Seidah N.G., Israel A.;
RT "The Notch1 receptor is cleaved constitutively by a furin-like convertase.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
RN [8]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).

[illegible]

Query Match	9.6%	Score 554;	DB 1;	Length 2531;
Best Local Similarity	21.7%;	Pred. No. 4.7e-21;		
Matches 309;	Conservative	33;	Mismatches 499;	Indels 582; Gaps 75
DOMAIN	1097	1143	EGF-LIKE 29.	CALCIUM-BINDING (POTENTIAL).
DOMAIN	1145	1161	EGF-LIKE 30.	CALCIUM-BINDING (POTENTIAL).
DOMAIN	1193	1219	EGF-LIKE 31.	CALCIUM-BINDING (POTENTIAL).
DOMAIN	1221	1265	EGF-LIKE 32.	CALCIUM-BINDING (POTENTIAL).
DOMAIN	1267	1305	EGF-LIKE 33.	
DOMAIN	1307	1346	EGF-LIKE 34.	
DOMAIN	1348	1384	EGF-LIKE 35.	
DOMAIN	1387	1426	EGF-LIKE 36.	
REPEAT	1445	1480	LIN/NOTCH 1.	
REPEAT	1481	1522	LIN/NOTCH 2.	
REPEAT	1523	1562	LIN/NOTCH 3.	
REPEAT	1517	1946	ANK 1.	
REPEAT	1950	1980	ANK 2.	
REPEAT	1984	2013	ANK 3.	
REPEAT	2017	2046	ANK 4.	
REPEAT	2050	2079	ANK 5.	
DOMAIN	1730	1733	POLY-ALA.	
DOMAIN	1891	1894	POLY-GLU.	
DOMAIN	2258	2261	POLY-PRO.	
DOMAIN	2497	2500	POLY-SER.	
SITE	1654	1655	CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY SIMILARITY).	
DISULFID	24	37	BY SIMILARITY.	
DISULFID	31	46	BY SIMILARITY.	
DISULFID	48	57	BY SIMILARITY.	
DISULFID	63	74	BY SIMILARITY.	
DISULFID	68	87	BY SIMILARITY.	
DISULFID	89	98	BY SIMILARITY.	
DISULFID	106	117	BY SIMILARITY.	
DISULFID	111	127	BY SIMILARITY.	
DISULFID	129	138	BY SIMILARITY.	
DISULFID	144	155	BY SIMILARITY.	
DISULFID	149	164	BY SIMILARITY.	
DISULFID	166	175	BY SIMILARITY.	
DISULFID	182	195	BY SIMILARITY.	
DISULFID	189	204	BY SIMILARITY.	
DISULFID	206	215	BY SIMILARITY.	
DISULFID	222	233	BY SIMILARITY.	
DISULFID	227	243	BY SIMILARITY.	
Query Match	9.6% <td>Score 554; <td>DB 1; <td>Length 2531;</td> </td></td>	Score 554; <td>DB 1; <td>Length 2531;</td> </td>	DB 1; <td>Length 2531;</td>	Length 2531;
Best Local Similarity	21.7%;	Pred. No. 4.7e-21;		
Matches 309;	Conservative	33;	Mismatches 499;	Indels 582; Gaps 75
28	CAAGAAATCATGAAGTCGCGAGTGGCAGACACCGC	-----CACCGCGC-----	68	
24	CSQPSGTCLNG-----GRCEVANGTEACVCSGAFVGQRCDPSPCLSPFCKNAGTCYVW	77		
69	-----GCCGC-----	-----GGGCCCCAGCTGCCGC-----TGNAGCGC	97	
78	DHGGIVDYACSCPLGFSGLCLTFLANACLIANPCRNNGTCDLLTLEYKRCPPPGWSGKS	137		
98	C-----GGCAAGACA-----	-----GGAGCGGTG-----GGCGAGGT-----	127	
138	CQQADPCASNPANGGOCLEFESSYTCGCPGPHGTCTQDVNECSQNFCLCRHGTCHN	197		
128	--GGTGGCGCTGTCTG---TCTGAGCAGAGCGTGCCG---	ATCTCGCGCTGCCGGGCG---	176	
198	EIGSVRCACRATHGPHCELPPVPCSPSPQNGGTCTPTGDTTACALPGFASQNCEN	257		
177	-----GCCGGCGCGCTGCG-----TGCCCT-----	-----GCTGGAC--GAG--CAGCAG	215	
258	VDCFPGNCKNKGACVDGVNTYNCRCPEWTGQVCTEDVDECOLMPNACQAGTCHNSHG	317		
216	GTAAACGTGCTGCTCT-----ACGACATGAAGCGTGTACTC-----	AC--	255	
318	GYNCVYNGWTGEDCSNIDDCASAACTGATCHDRVASFYCECPHGRGTGLLCHLNDA	377		
256	-----GCCCTAAG--GAGCTGGTCCG-----ACCTG--CCCCAGAC	290		
378	SNPCNEGSNCDTNPNVNGKAICTCPRGYTGPAQSQDVDECALGANPCBHAGKCLNTLGS	437		
291	CGCAGGTGACCA-----AGGTGGAGATTCTCCAG-----	CACG	324	

```
Db 628 SNPTFGVCRDINRYDCVQCPGFTGPLCNWEINEECASSPGCGGSGVDGNGFHCLCPP 687
QY 446 -----CTTGAC-----GG-----CGAGGCG-GCATCGGTTCTG-----C 475
Db 688 GSLPPLCLPANHPCAKPCSHGVCHDAPGPGFCQCDPWGSGFRCSQSLAPDACESQPCQA 747
QY 476 GGAC-----GATCGCA-----TCTGTGTGCG-496
Db 748 GGTCTSDGIGFHCTCAPGFGQHQCEVLSPTSLCEHGHCHCESDPDLQITVCSPPGWQGP 807
QY 497 -----CTGAAGC-----CCCT-----C-----CAGGAGCC 519
Db 808 RQOQVDVEGAGASPCGPHGTCTNLPSPFRCHGSGYTFPCDDIDCDPNPLNGSGCQ 867
QY 520 GCGGACCCGCCATCCAGGCGGCAAGAGGAATACGTCTCTG-TGGGTCTCCCCAA 578
Db 868 DGVGSPSCSLSGFA-----GPRCARDVDECLSPGPG-TCTDHVAFCTICPCPGYGG 920
QY 579 CGC-----CCCTCGCGGATCTGAGGGAGACA-----AGACGATCGCG-----GCCCA 623
Db 921 FHCETDLLPCSPSCFNGGTCVDGVNSFSCLCRPGYTGTHCYKVDFCRPCLHGGTCN 980
QY 624 CTCGCGCCT-TAACTG--CAT-----CCAGCCTGGGGCTGAGG-----658
Db 981 PTHSGPECTCREGFTGNQCNFVDWCSQAPCQNGRCVQTGAYCICPPENSGPLCDIPSL 1040
QY 659 -CTGAG--GCACTGGCAGG-----AGAGGCG-----GCTCCTC--T 690
Db 1041 PCTEAAHGVLEQLCQAGGCGCIDKDHSHYCVCPGSRMGSHCSQEVDPCTAQCQHGTT 1100
QY 691 CT-----GCACACTACTAGTC-----AC-----CAGAGACTTTAGGGGTGGGATTCACCT 737
Db 1101 CRGYMGVYCECTGYSGSDCEDDDVDECAQPCQNGGSCIDLK-----AHVLCSP 1151
QY 738 CGT-----GTGTT-----TC--TATTTTGAAGAAGC-----762
Db 1152 PGTGLVLSINEDDCGPGSLDGLRCLHNGTCVDLVGFCRCNPPGYTGLHCBADINEC 1211
QY 763 -AGACATTTTAAAAAATGTCACGTTTGGTG--CTTCTCAGATTTCTGAGGAATTCCTT 819
Db 1212 RPTCHAAHTRDCLQDPGGHFRICLPGFTGPRCQTALFPCEQPCQHGCGCRPSLG--1268
QY 820 TGTATTGTATATAC-----AATGACACCGACTGA 890
Db 1269 ----RGGLTFTCHCVQPPWGLRCSRVARSRELOCPVIGPCQQTARGPCACPPGLUGP 1324
QY 851 G---AATATTGTTTACAATA---GTTT-----TGTGGGCG-TGTTTTTT 889
Db 1325 SCRVSRAQSGATNTSCATPCLHGSCLPVQSVFFRCVCAPGCGGPRCETPSAAPEVP 1384
QY 890 GTATTAAACAA 901
Db 1385 EPRCPRAACQA 1396

RESULT 14
NTCI RAT
ID NTC1 RAT STANDARD; PRT; 2531 AA.
AC Q0708;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
```

```
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
RL development.";
RN Development 113:199-205(1991).
RP [2]
RA REVISIONS TO 1652-1653.
RA Weinmaster G.;
RN Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RX [3]
FUNCTION.
MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RN Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RL neural progenitor cells to an astroglial fate.";
RN Neuron 29:45-55(2001).
RX [4]
TISSUE SPECIFICITY.
MEDLINE=93202015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RN "Notch2: a second mammalian Notch gene.";
RL Development 116:931-941(1992).
RX [5]
TISSUE SPECIFICITY.
MEDLINE=21331789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RN "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT functional roles for the Notch-DSL signaling system during brain
RL development.";
RL J. Comp. Neurol. 436:167-181(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
CC Expressed in postnatal central nervous system (CNS) germinal zones
CC and, in early postnatal life, within numerous cells throughout the
CC CNS. Found in both subventricular and ventricular germinal zones.
CC -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
CC days 12 and 14 and decrease rapidly to much lower levels in the
CC adult.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 36 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```


or send an email to licenses@isb.sib.ch).

```

CC CC      EMBL; U97669; AAB91371.1; --
DR DR      EMBL; AF058900; AAC14346.1; --
DR DR      EMBL; AF058981; AAC14346.1; JOINED.
DR DR      EMBL; AF058982; AAC14346.1; JOINED.
DR DR      EMBL; AF058983; AAC14346.1; JOINED.
DR DR      EMBL; AF058984; AAC14346.1; JOINED.
DR DR      EMBL; AF058985; AAC14346.1; JOINED.
DR DR      EMBL; AF058986; AAC14346.1; JOINED.
DR DR      EMBL; AF058987; AAC14346.1; JOINED.
DR DR      EMBL; AF058988; AAC14346.1; JOINED.
DR DR      EMBL; AF058989; AAC14346.1; JOINED.
DR DR      EMBL; AF058990; AAC14346.1; JOINED.
DR DR      EMBL; AF058991; AAC14346.1; JOINED.
DR DR      EMBL; AF058992; AAC14346.1; JOINED.
DR DR      EMBL; AF058993; AAC14346.1; JOINED.
DR DR      EMBL; AF058994; AAC14346.1; JOINED.
DR DR      EMBL; AF058995; AAC14346.1; JOINED.
DR DR      EMBL; AF058996; AAC14346.1; JOINED.
DR DR      EMBL; AF058997; AAC14346.1; JOINED.
DR DR      EMBL; AF058998; AAC14346.1; JOINED.
DR DR      EMBL; AF058999; AAC14346.1; JOINED.
DR DR      EMBL; AC004257; AAC04897.1; --
DR DR      EMBL; AC004563; AAC15789.1; ALT_INIT.
DR DR      PIR; S78549; S78549.
DR DR      HSP; P00740; LEDM.
DR DR      Genew; HGNC:7883; NOTCH3.
DR DR      MIM; 600276; --
DR DR      MIM; 125310; --
DR DR      InterPro; IPR002110; ANK.
DR DR      InterPro; IPR000152; Asx hydroxyl_s.
DR DR      InterPro; IPR000742; EGF-2.
DR DR      InterPro; IPR001881; EGF-Ca.
DR DR      InterPro; IPR001438; EGF-II.
DR DR      InterPro; IPR006209; EGF-like.
DR DR      InterPro; IPR002049; Laminin_EGF.
DR DR      InterPro; IPR008297; Notch.
DR DR      InterPro; IPR000800; Notch_dom.
DR DR      Pfam; PF00023; ank; 6.
DR DR      Pfam; PF00008; EGF_34.
DR DR      Pfam; PF00066; notch; 3.
DR DR      PIRSF; PIRSF002279; Notch; 1.
DR DR      PRINTS; PR00010; EGFBL00D.
DR DR      PRINTS; PR00011; EGFBLAMIN.
DR DR      PRINTS; PR01452; NOTCH.
DR DR      SMART; SMO0248; ANK; 6.
DR DR      SMART; SMO0179; EGF_CA; 19.
DR DR      SMART; SMO0004; NL_3.
DR DR      PROSITE; PSS0297; ANK_REPEAT; 1.
DR DR      PROSITE; PSS0088; ANK_REPEAT; 4.
DR DR      PROSITE; PSS0010; ASX_HYDROXYL; 18.
DR DR      PROSITE; PSS0022; EGF_1; 33.
DR DR      PROSITE; PSS0186; EGF-2; 25.
DR DR      PROSITE; PSS0026; EGF-3; 34.
DR DR      PROSITE; PSS0187; EGF_CA; 16.
DR DR      Receptor; Transcription regulation; Activator; Differentiation;
KW      Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW      Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW      Disease mutation.
FT FT      SIGNAL 1 39
FT FT      CHAIN 40 2321
FT FT      CHAIN 1629 2321
FT FT
FT FT      CHAIN 1662 2321
FT FT
FT FT      DOMAIN 40 1643
FT FT      TRANSMEM 1644 1664
FT FT      DOMAIN 1665 2321
FT FT      DOMAIN 40 77
FT FT      DOMAIN 78 118
FT FT      DOMAIN 119 156
FT FT      DOMAIN 158 195

```

```

FT DOMAIN 197 234 EGF-LIKE 5.
FT DOMAIN 236 272 EGF-LIKE 6. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 274 312 EGF-LIKE 7.
FT DOMAIN 314 350 EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 351 389 EGF-LIKE 9.
FT DOMAIN 391 429 EGF-LIKE 10. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 431 467 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 469 505 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 507 543 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 545 580 EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 582 618 EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 620 655 EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 657 693 EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 695 730 EGF-LIKE 18.
FT DOMAIN 734 770 EGF-LIKE 19.
FT DOMAIN 771 808 EGF-LIKE 20.
FT DOMAIN 810 847 EGF-LIKE 21. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 849 885 EGF-LIKE 22. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 887 922 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).

Query Match 9.9%; Score 570.5; DB 1; Length 2321;
Best Local Similarity 23.6%; Pred. No. 6.7e-296; Indels 517; Gaps 79;
Matches 321; Conservative 26; Mismatches 496;

QY 20 CCAGTCGCCAAGAAATCATGAAAGT---CGCCAGTGG-----CAGCACCG-- 60
DB 82 CHSGPC---AGRGVQSSVVAGTARFSCPRGRGPDCLSLPDLSSPCAHCARCSVG 137
QY 61 -----CCAC-----CGCGCGCGCGGCCAGCTGCG--GCTGAA----- 94
DB 138 PDGRFLCSPFGYQGRSCRSDVDECRVGEPRHGHGTCINTPGSPRCPPAGYTGPLCENP 197
QY 95 -----GGCC--GGCAAGACAGCAGCGGTGCG-----GGCGCAGGT----- 127
DB 198 AVPCAPSPCRNGGTGROSGLDLYDCA-CLPGFEGQCNCEVNVDDCPGHRCLNGGTGCVDGVN 256
QY 128 -----GGTGGCGCT-----GTCCTCTGAGCAGACGGTG--GC----- 157
DB 257 TYNQCPPEWTGQFCTEDVDEBQCLQPNACHNGGTGCTNTLGHSCVCVNGMTGBSCSNID 316
QY 158 -CATCTC--GGCTGCGCGGGCGCG- CGCGGGCGCGCTGCTGCG----- 197
DB 317 DCATVCFHGATCHDRVASFYCACPMKGTGLLCHLDDACVSNPCHEDAICDTNPVNGRAI 376
QY 198 CT---GCTGGAC-----GAG-----CAGCAGTAAACGTGTGCTCTA----- 232
DB 377 CTCPPGFTGGACDQDVDECSIGANPCHELRGVNTQGSFLCQCGRGTGPRCETDVNECL 436
QY 233 CGACATGAAGCGCTGTTACTC-----ACGCTCAAGGAGC---TGCTGCCC 275
DB 437 SGPCRNQATCLDRIGQFTCTCMAGFTTYCEVDIDECQSSFCVNGGV-CKDRVNGFSCTC 495
QY 276 ACCCTGCCC-----CA-----GAAC-----CGCAAGGTGAGCAAG----- 305
DB 496 PSGFSGSTCQLDVDECASTPCRNAGAKVDQPDGVEYECRCAEGFETLCDRVNDDCSPPDCH 555
QY 306 ---GTGAGATTCTCCAGCAGGTATCATGATCATCAGGAC-----CTTCAGTTG- 353
DB 556 HGRVVDGIASFSCACAFYGTGTRCESQVDECRSPQCRHGGKCLDLVDKYLICRCPSGTGV 615
QY 354 -----GAGCTGAATCG-CAATCCGAAGTTGG-----GACCCC-CGGGGGCGCGA 395
DB 616 NCEVNIDDCASNPCPTFGVCRDGINRYDVCQPGFTGPLCNVNEICASSPCGEGSGVDG 675
QY 396 GGG-----CT--GCCGCTC--CGGGCTC--CGCTCAGCACCTCA- 429
DB 676 ENGRFCLCPGSLPPLCLPPSHPCAHPGSHGICYDAPGGFRVCRPFGWGSQSLAR 735
QY 430 -ACGG--CGAGATCAGCG-----CCCTGACGG--C-----CGAGCG--GGCATCGGTTC 472
DB 736 DACSQPCRAGGTCSSDGMGFHCTCTCPGVQGRQCELLSPCTNPCEHGGRCESAPGQLPV 795
QY 473 TGC--GGACGATC-----GCATCTTGTGTGCG-----CTGAAGCGCCTC-----CC----- 510

```


SQ	SEQUENCE	801 AA; 65407 MW; EA54C9BF45A00F41 CRC64;
	Query Match.	10.9%; Score 627; DB 1; Length 801;
	Best local Similarity	28.5%; Pred. NO. 4.7e-25;
	Matches	280; Conservative 24; Mismatches 337; Indels 342; Gaps 44;
QY	2	GGGCCCATCTGTTC-----AGCCAGT---CGCCAAGAATCATGAAATGCCAGTGG 52
DB	83	GGGAYAAAAAVSPLLAPINQFVATGRPLINGANGAP--GTGANGPGGLNGG 140
QY	53	CAGCACCGCCACCCGCGCGGCCCCAGCTGC CGCTGAAGCCCGCAAGACGC--- 109
DB	141	AGSGAFC---AGAGNGGAGG---LFGSGGAGASTDVAGGAGGAGGAGNAGMLFG 192
QY	110	GAGCGGTG-----CGGGCAGGTGG-----TGCGCTCTCTCTGACACAGCGTG 155
DB	193	AAGVGGVGFPSNGATGAGAGGAGGLFGAGRERGSGLITG---GAGAG---GNA 246
QY	156	GCCATCTCGCGCT-GCCGGGGCGCGGGCGCGCTGCTGCCCTGCTGGA-----CGA 208
DB	247	GTLATGDGAGGTGASRSGGFGGAGGAGBGMFFG-----SGSGGAGGISKSYVD 299
QY	209	GCAGCAGTAACGTGCTCTACGACATGAACGGCTGTACTCACGCTCAAGGAGC- 267
DB	300	SAAGGAGAPGLI NGNGNG---GNHGASTGGDGGPG-----AGGTGVL 341
QY	268	--TGTGCCCCCTTGCCCCAGAACCGGAGTAGAAGGTGAGATCTCCAGCACGT 325
DB	342	I GNGNG-----GSGGTGATLGKAGTGGTG-----GVT 368
QY	326	CATCGACTACATCA-----GGGACCTTCAGTTGAGCGTG 359
DB	369	LLGLDGFTAPASTSPHLTLQQDIVNWNPQTLTGRPLINGANGTFGTGADGAG--G 426
QY	360	AATCTGGAAATCGGAATGGAGACCCCGGGGCCGAGGCTGCCGCTCCGGCTCCGCTC 419
DB	427	WLFNGNGN---GGGGTTGG-----VNGGAGGAGGAGILFG-TGGT---GSGSGPGATG 473
QY	420	AGCACCTTCAACGGCGAGATCAGCGCCCTGACGGCGGAGGGGATGCTTCTCGCGAC 479
DB	474	LG-----GIGGAGGAALLFG-----SGGAGSSGAGAVGNG-----GAGNA 511
QY	480	GATCGCATCTGTGCTGAAGCGCTCCCAGGACC GGCGGACCCAGCCATCCAG 539
DB	512	GALLGAA-----GAGGAGGAG-----AVGNN---GGAGN---GGLFANGAG 548
QY	540	GGGCAAGAGGAAATTACGTGCTCTGTGGGTCTCCCCCAACGGCCCTCGCCGATCTGAGG 599
DB	543	GPFGPFSFAGAGGIGGAGNGGLFGAGG-----TGGAGGSGSTLAGAG 591
QY	600	GAGAACAGACCGATCGCGGCCACTGCGCCCTTAACCTGATCAGCCTGGGCTGAGGC 659
DB	592	GAG---GNGGLFGA--GGTG-----GAGSHSTAAGVSG-----GAGGA 624
QY	660	TGAGCACTGCGAGGAGGCGGCTCTCTGTGCACACCTACTAGTACCAGAGACTTT 719
DB	625	GDA LLSLG--ASGGAGSGGSSLT-----AAGVVGGIGGAGGLLFG 665
QY	720	AGGGGTGGGATTCACCTCGTGTCTTCTATTTTTTGAAGCAGACATTTTAAAATG 779
DB	666	SGGAGSGG-----FNSNGGAGGAGGDLGLVSGGAGGA 702
QY	780	GTCAGGTTGGTGCTCTCAGATTTCTGAGGAATTTGCTTTGTATGTATATCAATGA 839
DB	703	GASATGAATGDDG-----GAGGKSAGF-----GAGGKSGAF----- 725
QY	840	TCACCGACTGAGATATGTTTACAATAGTCT--GTGGGCTGTTTTTTTGTATTAA 897
DB	726	----LGGDGGAGGATGLSGATHGKGVGVS AVLI GNGNGNG-----GNSGN----- 771
QY	898	ACAAATAATTTAGATGGTGA AAA 920
DB	772	-----AGKSGGAPGPSGAGA 787

```

RESULT 11
PHX5 MOUSE STANDARD; PRT; 672 AA.
ID PHX5 MOUSE AC
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Par-hexamer repeat protein 5.
GN PHX5 OR PER.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM M.A.
RX MEDLINE=86014384; PubMed=2413365;
RA Shin H.S., Bargiello T.A., Clark B.T., Jackson F.R., Young M.W.;
RT "An unusual coding sequence from a Drosophila clock gene is conserved
in vertebrates.";
RL Nature 317:445-448 (1995).
CC -|- SIMILARITY: TO OTHER G-T STRETCH CONTAINING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M12039; AAA88320.1; -
DR EMBL; X02966; CAA26710.2; -
DR PIR; A24403; UMMS
DR MGD; MGI:104521; Phx5.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR Repeat.
FT DOMAIN 59 672 G-T REPEATS.
SQ SEQUENCE 672 AA; 57924 MW; E855FF428CF424C0B CRC64;

Query Match 10.7%; Score 618.5; DB 1; Length 672;
Best Local Similarity 31.4%; Pred. No. 1.1e-24;
Matches 266; Conservative 18; Mismatches 321; Indels 243; Gaps 39;

Qy 117 GCGGGCGAGGTGTGGCG-----TGT 137
Db 6 GCGHCKEAPRGSVACECPGFLAKNOKDICKYELSISEVGTAYTEVRKTTTGTATGI 65

Qy 138 CTGTCGTAG---CAGAGCGTGGCCATCTCCGCGTCGCGGGCGCGGGCGCGCGCTGCC 194
Db 66 ATGTCGTGRVATGRGTGTDTDGTVTARATVTARVTGTGTGTATVTETGTAKVTDGT 125

Qy 195 GCCTCTGTGCAGCAGCAGCAG-GTAAACGTGCTGCTACGACATGAACGGCTGTACTC 253
Db 126 G-----TGAKVTGTAKVTGTGTGTGTG-TG-----TGTGTGTGTGAKVT- 167

Qy 254 AGCCCTCAGAGAGCTGGTGTCCACCCCTGCCCAGAACCCAGAGGTAGCAG- GTGGAGA 312
Db 168 -----GTGTDRGTG-----TGT-----GTGTGTGTGTGTGTAKV 196

Qy 313 TTCTCCAGCAGCGTCATCGACTACATCAGGGACCTTCAGTTGAGCTGAACCTCGGAATCG 372
Db 197 T-----GTAKVTGTGTAKVTGTG-----TGTGTGTGTG-TGT-----GTDGTG 236

Qy 373 AGTTGGAGCCCCGGGGGCGCAGGGCTCGCGTCCGGCTCCGGCTCCGCTCAGCACCTCAAG 432
Db 237 TAKVTGTGT-----GTGTG-TGTGTGTGTG-----TGTAKVTGTGTD 275

Qy 433 GCGAGATCAGCGCCCTGACGGCGCGAGCGGGCATGGCTTCCTCGGACGATCGCATCT-TG 491

```

Tuberculin: RV2634C: -

InterPro: IPR000884; PE_region.
Pfam: PF00934; PE; 1.
ProDom: PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;

Query Match 10.9%; Score 629; DB 1; Length 778;
Best Local Similarity 31.4%; Pred. No. 3.7e-25;
Matches 268; Conservative 8; Mismatches 325; Indels 252; Gaps 37;

QY 37 TGAAGTCCAGTGCACACCGCCACCGCCGCGCGCGCCAGCTGCGCGCTGAAG- 95
DB 128 TGAAGTCCAGTGCACACCGCCACCGCCGCGCGCGCCAGCTGCGCGCTGAAG- 174
QY 96 GCCGCAACACACCGA---GCCGT-GCGGGGA---GGTGTGCGCTGTC-TGCTGAGC 147
DB 175 GLVSGAAGVGGRGGLLNGNGTGGAGAGATLVGTGGV-GATGLGSGFGGAGG 233
QY 148 AGAGCTGCGCATCTCGCGCTCCCGGGG-GCCGGGGCGCGCTCCCTGCGCTGCTGAC 206
DB 234 AAGVGTG-----GVGSGGVGVFGNG-----FGAG 263
QY 207 GAGCAGAGTAACCTGCTCTACGACATGAACGCTTTACTACGCTCAAGCTCAAGGAG 266
DB 264 GLGAAGVGGAAYFTGGG-----GVGSDGAP-GDGGAGPLLING-GVGGIGGAG 315
QY 267 CTGTGTCACACC--TGCCCCAGAACCGCAAGTGAAG-----CAAGTGGAGATTCT 316
DB 316 AAGNGGAGGMLLDGAGGGGQGPVAGVLGMPGAGGNGNANWFGSGAGGGGCT--- 372
QY 317 CCAGCAGCTATCGACTATCATCAGGACCTTCAGT---TGGAGCTGACTCGGAATCGA 373
DB 373 -----GLAGTNGVNPFSIANPNTGANGTDSNGNTGNGGPGPA---GGVGEAGV 422
QY 374 AGTTGGACCCCGGGGGCGGAG--GCGCGCGCTCCGGCTCCGCTCAGCACCTCAAC 431
DB 423 GGGGLGSLDNDGTGKGAGGTAGTDGAG-----AG 458
QY 432 GCGCAGATCAGCCCTGACGCGCGAGGCGGATCGTTCCTCGGAGCATCGCATCTTG 491
DB 459 GAGGIGTDSAGVATGEGSDGATGGVGGVAG-----GKGQ-GHNTGVGDAPGG 512
QY 492 TGTGCTGAAGCCCTCCCGCAGGACCGCGGAGCCAGCCATCCAGGGGCGCAAGAGGA 551
DB 513 DGGIGDNGALG-----AAGNGGTGAGG-----NGRGMGLIGNGA 552
QY 552 ATTACGTGCTCTGTGGTCTCCCCAACCGGCTCCCGGATCTCAGGAGAAACAAGACC 611
DB 553 G-GAGGTG-----GTGGG-----GAAGFAGGVGAGGEGLTGDAGTAE 589
QY 612 GATCGCGGCACTGCGCCCTTAAGTGCATCCAGCTCGGGGCTGAGGCTGAGGCACT--- 668
DB 590 GGT-GGLGG-----LGVGGTGMGSGGSGVGGNGGAGSLIG 625
QY 669 --GCGCAGGAGAGGGGCGCTCCTCTGTCACACTACTACACAGACTTTAGGGGCT 726
DB 626 LGGCGGAGVGGTGGIG-----GIGG-----AGGNGGA 653
QY 727 GG-GATTCCACTGCTGTGTTCTATTTTGAAGACACAGATTTTAAAAATGTCACG 785
DB 654 GGATTTGGGATGGGGT-----GGVGGAGCTGGTGGAGGTTGSGGAG 698
QY 786 TTTGGTCTCTCAGATTTCTGAGGAAATGCTTTGTATGTATATATCAATGATCACC 845
DB 699 GLIGWAGAGGTGAGTGGGGGGLGGGGGNG--NGGTGATGGGDFALGGNG----- 749
QY 846 ACTGAGATATG 858

Tuberculin; RV2634C; -
InterPro: IPR000864; PE_region.
Pfam: PF00934; PE; 1.
DR ProDom: PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 V -> L (IN REF. 2).
FT CONFLICT 63 Q -> H (IN REF. 2).
FT CONFLICT 74 A -> T (IN REF. 2).
SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;
Query Match 10.9%; Score 629; DB 1; Length 778;
Best Local Similarity 31.4%; Pred. No. 3.7e-25;
Matches 268; Conservative 8; Mismatches 325; Indels 252; Gaps 37;

QY 37 TGAAAGTCGCCAGTGCACACCGCCACCGCGCGGCCCCAGCTGCGCGTGGAAG- 95
DB 128 TGAPGPGLLNGENGNG-----SGAPGPFGAGDAGLIGNG-GTGGKGD 174
QY 96 GCCGCAACACACGCA---GCCGT-GCGGGGA---GGTGTGCGCTGTC-TGCTGAGC 147
DB 175 GLVGSAGAGVGGRGLWLLNGTGGAGAATLVGTGGVG-GATGLIGSGFPGAGG 233
QY 148 AGAGCGTGCCATCTCGCGCTCCCGGGG-GCCGGGGCGCGCTCCCTGCGCTGCTGAC 206
DB 234 AAGAVTTG-----GVGSGGVGVFGNG-----FCGAG 263
QY 207 GAGCAGAGTAACCTGCTCTACGACATGAACGCTTTACTACGCTCAAGCTCAAGGAG 266
DB 264 GLGAAGVGGAAYFTGGG-----GVGSDGAP-GDGAGGPLLIONG-GVGGIAGG 315
QY 267 CTGTGTCACCACC--TGCCCCAAGAACCGCAAGTCAAG-----CAAGTGGAGATTCT 316
DB 316 AAGNGGAGGMLLDGAGGGQGPAVAGVLGMPPGAGGNGNANWFGSGAGGGGCT--- 372
QY 317 CCAGCAGCATCGCATCATCAGGACCTTCAGT---TGGAGCTGACTCGGAATCCGA 373
DB 373 -----GLAGTGNVFNPSIANPTANGTDSNGNTGNGNGPGPA---GGVGEAGVV 422
QY 374 AGTTGGACCCCAGGGGGCGGAG--GCGCGCGCTCCGGCTCCGCTCAGCACCTCAAC 431
DB 423 GGCGLGESLDNDGTGKGAGTAGTDGAG-----AG 458
QY 432 GCGCAGATCAGCCCTGACGCGCGAGGCGGATCGTTCCTCGGAGCATCGCATCTTG 491
DB 459 GAGGIGTDSAGGATVATGEGSDGATGGVDGAG-----GKGGQ-GHNTGVGDAPGG 512
QY 492 TGTGCTGAAGCCCTCCCGCAGGACCGCGGAGCCAGCCATCCAGGGGCGCAAGAGGA 551
DB 513 DGGTGGDNGALG-----AAGNGGTGAGG-----NGRGGMLIGNGA 552
QY 552 ATTACGTGCTCTGTGGTCTCCCCAACCGGCTCCCGGATCTCAGGAGAAACAAGACC 611
DB 553 G-GAGGTG----GTGGG-----GAAGFAGGVGAGGEBLTGDAGTAE 589
QY 612 GATCGCGGCACTCGCCCTTAAGTGCATCCAGCTCGGGGCTGAGGCTGAGGCACT--- 668
DB 590 GGT-GGLGG-----LGGVGTGCMGSSGVGNGGAAGSLIG 625
QY 669 --GCGCAGGAGAGGGGCGCTCCTCTGTCACACTACTACACAGACTTTAGGGGCT 726
DB 626 LGCGGAGGVGGTGGIG-----GIGG-----AGGNGGA 653
QY 727 GG-CATTCCACTCGTGTGTTCTATTTTTTGAAGACACACATTTAAAAAATGTCACG 785
DB 654 GGAGTTGGGANTGGGGT-----GGVGGAGCTGGTGGAGGTTGSSGAG 698
QY 786 TTTCGTCTCTCAGATTTCTGAGGAAATGCTTTGTATGTATATATCAATGATCACC 845
DB 699 GLIGWAGAGGTGAGTGGGGGLGGGGGNG--NGGTGATGGGDFALGNG----- 749
QY 846 ACTGAGATATTG 858

Tuberculin; RV2634C; -
InterPro: IPR000884; PE_region.
Pfam: PF00934; PE; 1.
DR ProDom: PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 V -> L (IN REF. 2).
FT CONFLICT 63 Q -> H (IN REF. 2).
FT CONFLICT 74 A -> T (IN REF. 2).
SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;
Query Match 10.9%; Score 629; DB 1; Length 778;
Best Local Similarity 31.4%; Pred. No. 3.7e-25;
Matches 268; Conservative 8; Mismatches 325; Indels 252; Gaps 37;

QY 37 TGAAAGTCGCCAGTGCACACCGCCACCGCGCGCGGCCCCAGCTGCGCGCTGAAG- 95
DB 128 TGAPGPGLLNGENGNG-----SGAPGPFGAGDAGLIGNG-GTGKGGD 174
QY 96 GCCGCAACACACGCA---GCCGT-GCGGGGA---GGTGTGCGCTGTC-TGCTGAGC 147
DB 175 GLVGSAGAGVGGRGLWLLNGTGGAGAATLVGTGVGV-GATGLIGSGFPGAGG 233
QY 148 AGAGCGTGCCATCTCGCGCTCCCGGGG-GCCGGGGCGCGCTCCCTGCCCTGCTGGAC 206
DB 234 AAGAVTTG-----GVGSGGVGVFGNG-----FCGAG 263
QY 207 GAGCAGAGTAACCTGCTCTACGACATGAACGCTTTACTACGCTCAACCTCAAGGAG 266
DB 264 GLGAAGVGGAAYFTGGG-----GVGDGAP-GDGAGGPLLIONG-GVGGIAGG 315
QY 267 CTGTGTCACCACC--TGCCCCAAGAACCGCAAGTCAAG-----CAAGTGGAGATTCT 316
DB 316 AAGNGGAGGMLLDGAGGGQGPAVAGVLGMPPGAGGNGNANWFGSGAGGGCGT--- 372
QY 317 CCAGCAGCATCGACTATCATCAGGACCTTCAGT---TGGAGCTGACTCGGAATCCGA 373
DB 373 -----GLAGTGNVFNPSIANPTANGTDSNGNTGNGNGPSPA---GGVGEAGVV 422
QY 374 AGTTGGACCCCAGGGGGCGGAG--GCGCGCGCTCCGGCTCCGCTCAGCACCTCAAC 431
DB 423 GGCGLGESLDNDGTGKGAGTAGTDGAG-----AG 458
QY 432 GCGCAGATCAGCCCTGACGCGCGAGGCGGATCGTTCCTCGGAGCATCGCATCTTG 491
DB 459 GAGGIGTDSAGVATGEGSDGATGGVDGVGAG-----GKGGQ-GHNTGVGDAPFG 512
QY 492 TGTGCTGAAGCCCTCCCGCAGGACCGCGGAGCCAGCCATCCAGGGGCGCAAGAGGA 551
DB 513 DGGTGGDNGALG-----AAGNGGTGAGG-----NGRGGMLIGNGA 552
QY 552 ATTACGTGCTCTGTGGTCTCCCCAACCGGCTCCCGGATCTCAGGAGAAACAAGACC 611
DB 553 G-GAGGTG----GTGGG-----GAAGFAGGVGAGGEBLTGDAGTAE 589
QY 612 GATCGCGGCACTCGCCCTTAACCTGATCCAGCTCGGGGCTGAGGCTGAGGCACT--- 668
DB 590 GGT-GGLGG-----LGGVGTGCMGSSGVGNGGAAGSLIG 625
QY 669 --GCGCAGGAGAGGGGCGCTCCTCTGTCACACTACTACACAGACTTTAGGGGCT 726
DB 626 LGCGGAGGVGGTGGIG-----GIGG-----AGGNGGA 653
QY 727 GG-CATTCCACTCGTGTGTTCTATTTTTTGAAGACACACATTTAAAAAATGTCACG 785
DB 654 GGAGTTGGGANTGGGGT-----GGVGGAGCTGGTGGAGGTTGSSGAG 698
QY 786 TTTCGTCTCTCAGATTTCTGAGGAAATGCTTTGTATGTATATATCAATGATCACC 845
DB 699 GLIGWAGAGGTGAGTGGGGGLGGGGGNG--NGGTGATGGGGPALGNG----- 749
QY 846 ACTGAGATATG 858

```
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 665 V -> L (IN REF. 1).
FT CONFLICT 662 662 S -> T (IN REF. 1).
FT CONFLICT 672 672 NPGLSGCDVLIOALLEVVSALIOILGSSSIGOVNYSAGQA
FT CONFLICT 695 747 TQIVGQSVYQAL -> ILVFLDWSSFKLSRLFLLRS
      (IN REF. 1).
SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;

Query Match 11.0%; Score 635.5; DB 1; Length 747;
Best Local Similarity 30.6%; Pred. No. 1.7e-25;
Matches 280; Conservative 21; Mismatches 358; Indels 255; Gaps 39;

Qy 30 AGATCATGAA-----AGTGCACGTGGCAGCACCACCGCGCGCGCGCCAGCT 84
Db 3 AGAAAAAGAGGQGGYGLGGQAGQGGYGLGGQ-----AGGAGAAAAAAG 52
Qy 85 GC-----GGCTGAAGCGCGCAAGACAGCAGCGTGTGGG-----GCGAGTGTGTGCGCTGT 137
Db 53 GAGGGYGLGSGQAGRGQGGAGAAAAAGAGAGGQGGYGLGSGQAGRGGLGGQAG-----108
Qy 138 CTGCTGAGCAGAGCGTGGCCATCTCCGCTGCGCGGCGCGCGCGCGCGCTGCTGCC 197
Db 109 -----AAAAAAGAGAGQGGYGLGNQAG-----RGQGAAAAAAGAGAGQGGYGLGSGQGA 159
Qy 198 CTGCTGACAGCAGCAGGTAAAGCTGCTGTCTTACACATGAACGCTCTTACTCAQGC 257
Db 160 GRGGLGQ-GAGAAAAAGAGGQGGY-----GLGGQAGQGGYGLGSG-----202
Qy 258 CTCAGAGAGTGTGCCACCCCTCCCGCAGACCGCAGTGTAGCAGGTGGAGATCTC 317
Db 203 ---CGAGRGGLGGQ-----AGAA-AAAAAGAG-----QGGGLGGQ-----235
Qy 318 CAGCAGCTCATCGACTATCAGGAGACCTTCAGTTGAGCTGAACTCGGAATCCGAAGTT 377
Db 236 -AGGAG-----ASAAAAGAG-----QGGYGLGSGQAGRGGAGAAAAAGGA 280
Qy 378 GGGACCCCGCGCGCGCAGGCTCCCGTCCGGCTCCGGCTCCGCTCAGCACCTCAACGCGAG 437
Db 281 GGGY-----GGLGGQAGQGGYGLGSGQAGRG-----GLGGQAG 317
Qy 438 ATCAGCGCTGACGCGCGGCGCATGCTGCTCTCGGAGCATGCTGCTGTGTCG 497
Db 318 AAAAG-----GA-----GQGLGGQAGQ-----AGAAAAAGGA-----GQGYGG 355
Qy 498 TGAAGCGCTCCCGCAGGAGCGCGGACCCAGCCATCCAGGCGGCAAGAGAAATACG 557
Db 356 LGSQAG-----RGGLGQ-----GAGAVAAAAAGGAG-----QGGYGLGSGQAG 396
Qy 558 TGCTGTGGTCTCCCCACGCGCTCCGCG--GATCTGAGG-GAGACACAGCCGAT 614
Db 397 RG-----GQAG-----AAAAAGGAGQGGYGLGNQAGRGGLGGQAGAA-AAAAAGGAG 447
Qy 615 CGGCGGCACCTCGCGCCCTTAATCATCATCCAGCTCGGCGCTGAGCCTGAGCAGCTGCGAG 674
Db 448 QGGYGG-LGNQAG-----RGQGAAGAGAGQGG-----YGLGSGQ 484
Qy 675 GAGAGGCGCTCTCTGACACCTACTA-----GTCACGAGACTTTAGGGGTGG 728
Db 485 GAGRG-----QGAGAAAAAAGVAGQGGIRGQAGQGGYGLGSGQSGRGGLG 533
Qy 729 GATTCACCTGCTGTGTTCTATTTTAAAAAGCAGACATTTAAAAAATGTCACGTTT 788
Db 534 QAGAAAAAAG-----GAGQGGGQAGGAGGAGAAAAAGGVRQGG--Y 575
Qy 789 GGTGCTCTCAGATTCAGAGAAATGCTTTGTTATTGTTATATATATATATATATATATAT 848
Db 576 GGLG-----SQAGRGQGGAGAAA-----AAGGAGQGGYGLG 609
Qy 849 GAGATATGTTTACATAGTCTGTGCGGCTGTTTTTTTGTATTATAAACAATAATTT 908
Db 610 GQGVGRGLGGGAGAAAGAGAGQGGYGGV-GSASASASAAARLSFPQSSRVSSAVSN 668
```

QY 909 AGATGGTCARAAA 922
Db 669 LVASGPTNSAALSS 682

RESULT 9

YQ34_MVCTU STANDARD; PRT; 778 AA.
AC F71933;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PE-PGRS family protein RV2634C/MT2712/MB2667C.
GN RV2634C OR MT2712 OR MTCY441.04C OR MB2667C.
OS Mycobacterium tuberculosis, and
OC Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
RT Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";
RT J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H., Pryor M., Duthey S., Grondin S., Lacroix C., Monsenpe C., Simon S., Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R., Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z80225; CAB02341.1; -
CC EMBL; A8007103; AKA47026.1; ALT INIT.
CC EMBL; BX248343; CAD94852.1; -
CC PIR; F70963; F70963.
CC TIGR; MT2712; -

[illegible]

RESULT 8
SPD1 NEPCL

QY 151 GCGTGGCCATCTCGCGCTGCGGGGGC---GCCGGGGCGCGCTGCTGCGCC---TGCTGT 203
 Db 493 TCENSLVLAICDXDTSRACINWKGKQCVLAS-----SATTTHADCTYHSTCTL 545
 QY 204 GACGAGCA---GCAGGT-AAAGCTGCTCTTACGACATGAACGCG--TGTTACTCAG 256
 Db 546 SNGTGCVPPLKCEALITTEACNLKANGOPCGWNGSQCIDKACSTASKTFTTTTSCQIGH 605
 QY 257 CTTCAAGGAGCTGGT---GC-----CCACCTGCG-----CCAGAA 289
 Db 606 ISTCVANNPVTNNGSLTIQCDPLTSCAARKSENCEIARVGFPTCLWVSSSTSCVEKS 665
 QY 290 CCGCAAGGTGAGCAAGTGGAGATTC--TCCA---GC-----A 322
 Db 666 CATASTVGTGALSAGGTFPGGCTYINTCISNNTADGCIAPSSGSSILVSNCRDGSKA 725
 QY 323 CGTCATCGACTACATCA---GGGAC-----CTT-----CA 349
 Db 726 SGDCYWNGSCVDKTCANITLTSHASCYSIFNOCTVNNGTACTQTLATATSYSTQENCK 785
 QY 350 GTTGGAGC--TGAACCTCGAATCGAAGTTGGACCCCG--GGGGCC 393
 Db 786 FTSNKNVWVGLACR---NATCADAPDITAYDSDECLAYPTPSETCTVVKVAGQGV 842
 QY 394 GAGGCGC-----TGCGGTCGCGGCTCC-- 415
 Db 843 SKSANCSDYMTSAQCHKLTNLNTANDCKWIVDRCVALSFFATGAC--TTFKGNKTMCRG 900
 QY 416 ---GCT---CAGCACCTCAACGGCAGATCAGC---GCCCTGAGC-GC---CG 456
 Db 901 YRAGCTNVTGAASSASCTLDCTLKT---GSLTFADCOALDSTCVSKXDGTCIVIGST 956
 QY 457 AGCGGCGATCGTCTCTCGGAGCATCGC---ATCTGTGTGTC---CT 498
 Db 957 CAGVGSFATNCFRSSASGTAGYCAMNTNCSQVTSAAECFAVTGLTGLDHSKQLYHSSCT 1016
 QY 499 GAA-GGCG---CTCCCCAGGAGCGCGGAC---CC---CA---GCCATCCAG 539
 Db 1017 SLKDGTCQBYKTACSSYATGNTCANSVQKGFDDATDCLRFANCASITGTGLTNTICVT 1076
 QY 540 GGGCAAGAGGAATTAGC-----TGCTCTGTGGGTC---TCCC-----C 575
 Db 1077 YDFCVANVG---TAQCKLATCAAVLTQNSCSTSTAGTCAMSGSACLTVDVANVATEC 1133
 QY 576 CAACGCGCCTCGCGG-GATCTGAGGAG---AACAGACCGATCG-GGGGCCACTGGC 628
 Db 1134 AYITGTGLTNAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCSQSDAGLCWSGSA 1193
 QY 629 C-----CCTTAACCTGC-----ATCC 643
 Db 1194 CLTVVDANVATECPYITGTGLTNAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCS 1253
 QY 644 ---AGCCT-GGGGCTGAGGCTGAGGCACTCGCGAGGAGGGCGCTCCTCTCTGCAC--- 696
 Db 1254 QSDAGLCAWSGSACLTVDANVATECPYITGTGLTDAICAG--YNKACTVNRAGTACQK 1311
 QY 597 -----ACTACTAGTACCAGAGACTTTAGGGGTGGGATTCACCTGCTGTT 745
 Db 1312 EALCATYAAVQATCSQSDAGLCW--SGSACLTV-----DANVATECPYITGTGLT 1361
 QY 746 T-----CTATTT-----TTTCAAAAGC---AGACA-----T 768
 Db 1362 NAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCSQSDAGLCAWSGSACLTVDANV 1421
 QY 769 TTTAAAAATGGT---CA-----CGTTGGTGC-----TTCAGATT----- 803
 Db 1422 ATECAYITGTGLTDAICAGYNAKCTNLKDGTCODEKATCKLVTQNKCTSQITGLPGLSCL 1481
 QY 804 -----TCTG-----AGGAAATTTGCTT-----TGATTG 826
 Db 1482 WFDNSCPIITDVTCSAIVQSLDHAQCAQVSTGCTSVSDGSKQDFKTTCEQVAGTALSC 1541
 QY 827 TATATTACATGATC-ACCGACTGAGAAATATTGTTT-----TACA----- 865

Db 1542 TKTATSKCYLQGSNCITISNVATDCAKITSGATITVEICQSYNTGCVNARSACVQQQ 1601
 QY 866 -----ATAGTCTGTGGGCTGTTTTTTTGTATTAAACAATAATTTAGATGGTGA 919
 Db 1602 AQCSGYTSAMTSCYKSGAGLCIASNTDT-----ACVAATAATTCDAVILGTGNY 1652
 QY 920 AA 921
 Db 1653 SA 1654

RESULT 7

Y278 MYCTU STANDARD; PRT; 957 AA.
 AC P56877;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PE-PGRS family protein Rv0278c/MT0291 precursor.
 GN Rv0278C OR MT0291 OR MV035.06C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RP SEQUENCE FROM N.A.
 RN (1)
 RC STRAIN=H37Rv;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.,
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."; Nature 393:537-544 (1998).
 RL (2)
 RN SEQUENCE FROM N.A.
 RP STRAIN=CDC 1551 / Oshkosh;
 RC MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."; J. Bacteriol. 184:5479-5490 (2002).
 RL J. Bacteriol. 184:5479-5490 (2002).
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
 CC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AL021930; CAAL7353.1;
 DR EMBL; AE006936; AAK44511.1; ALT_INIT.
 DR PIR; D70835; D70835.
 DR TIGR; MT0291; --
 DR Tuberculist; Rv0278c; --
 DR InterPro; IPR000084; PE_region.
 DR Pfam; PF00934; PE; 1.
 KW Hypothetical protein; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 957 HYPOTHETICAL PE-PGRS FAMILY PROTEIN

1238 KAGTACQKATCNLYTTEATCTSTAAAAATADKCAWSGAACLAATVTTATECAVVTGTLT 1297
 299 GAGCAAGTGGAGATTCTC-CAGCAC-----GTC-----ATCGACTACAT--CAGGGAC 344
 1298 NAICAAYN-----ANCANKAGTACQKATCNLYTTEATCTSTAAAAATADKCAWSGA 1351
 345 C-----TTCAGTTG-----GAGCT-----GAAC-----TCGGAA 368
 1352 CLAATVTTATECAVVTGTLTNAICAAVYANNTANKAGTACQKATCNLYTTEATCTSTAA 1411
 369 TCCGAAGTGGGACC-----CCGGGG-----CCGAGGC-----TGC----- 402
 1412 AAATADKCAWSGAACLAATVTTATECAVVTGTLTNAICAAVYANNTANKAGTACQKATCNLYTTEATCTSTAA 1471
 403 CGTCCGGGCTCCGCTCAGC-----ACC-----CTCAACGGCG-AGATCAGCGCCCTGACG 452
 1472 CKDYTTTSNKTAQTSTLSCLWDNSVPTDLNCSVITGLGFVHAQCAQAYSNGCTSVSD 1531
 453 GCGGAGGC-----GGCATCGTTCTCGC---GGAGCATCGCAT-CTTGTCGCG-- 496
 1532 G-----SKQDFKSTCEQPGTTLGCTKASTKCVLOSAGITTSNVATDCAKITSGAGTI 1587
 497 ----CTG-AAGCGCTCCCCCAGGAGCGGCGGACCC-----AGCC-----ATCCAGG 540
 1588 TFEICQSYNTGCSVNARSACVQQAQCSGYTSAMTSCYKSGAGLCTASTNTDTACVAAT 1647
 541 GGGCAA---GAGGAATTAG---TGCTCTGTGGTCTCCCAACAGCGGCTCCGCGGA-T 593
 1648 AATCDVVLGAGNYSANENKAGCTNGT--TACVAKTCANAAGITFNHNTNCSYNT 1705
 594 CTGAGGGAGAACAGACCGATCGCG---GGCCACTCGCG---CCTTAATCGATCCAGC 646
 1706 CTVNSGNACQTMASKADQATQASCLYSVEGECVVGTSVVRKTCDTAATDTRDDTTC 1765
 647 CTGGGGCTGA-GGCTGAGGCA-----CTGG-----CGAGGA-- 676
 1766 STYQSCIVLALGACQARAAATYKSSLOCKFNTSGKCFWNTKTCVDLNCNLEAT 1825
 677 -GAGGGCGCTCCTCTCT---GC-----ACA-----CCTACTAGTC----- 707
 1826 LYDTHNECVAVDNLACTVTRATNGAAGAGCGMARGACASYTIBEQCKTNASNGVCVWNTNA 1885
 708 ----AC-----CAGAGACTTT-----AGGGGGTGGGATTCAC 736
 1886 NLPAACQDKSCTSAPTSTTTTHNDYAYNTATVKCTVATPSNSGNGPTLGGCQQTAA 1945
 737 -----TCGTGTGT-----TCTTA--TTTTTGAATA 760
 1946 SSYIDKEQOINANGPCWNGTQCADKSCATASATADYDDTKRAYITNKCTVSDSQ 2005
 761 GC-----AG-----ACATTT--TAAAAAATGGTC---ACGTTT 788
 2006 GCVEIPATCEMTQKQCYNKGADPCYWTGTACITKSCDNPADATATADECNLYLAGCTL 2065
 789 GGTGC-----TTCTCAGATTCTGAGGAATGCTTTGTATTGTATTATACAT 837
 2066 NNKCKTKVCEDFAPATDALCKQAISTCTTNG-----TNCVTRGTQFQALSGQCVTSST 2120
 838 GATCACCGA-----CTGAGAATAATGTTTTTACAATAGTTCTGTGGGGC-TGT 883
 2121 NQCEWIPAVLNASNVITSPAVCTIKNGSTAPILTSEACAGVFNCTTNGGGCVTKS 2180
 884 T-----TTTTGT-----TATTAACAAATAATTAGATCG 914
 2181 TCSAVTIDVACTALNGTVCAWDSANQKCRDKOCQDFSGTTHAACQA-----ORAGCTAG 2235
 915 TGAATA 920
 2236 AGKCA 2241

RESULT 6

G168 PARPR STANDARD; PRT; 2704 AA.
 ID G168 PARPR STANDARD; PRT; 2704 AA.
 AC P17053;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE G surface protein, allelic form 168 precursor.
 GN 1688.
 OS Paramesium primaurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Paramesium.
 OX NCBI_TaxID=5886;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=90172419; PubMed=2308165;
 RA Prat A.;
 RT "Conserved sequences flank variable tandem repeats in two alleles of
 RL the G surface protein of Paramesium primaurelia.";
 RL J. Mol. Biol. 211:521-535(1990).
 CC -!- FUNCTION: This protein is the surface antigen or immobilization
 CC antigen of Paramesium primaurelia.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DOMAIN: It has internal homologies and a highly periodic structure
 CC with 37 periods of about 75 residues, each period containing 8
 CC cysteines, except for four half periods. A variable part of 475
 CC residues comprises 4 almost identical periods in the middle of the
 CC protein.
 CC -!- MISCELLANEOUS: Expression of G protein occurs at low temperatures
 CC (14-32 degrees Celsius).
 CC -!- SIMILARITY: Contains 34 PSA repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X52133; CAA36378.1; -.
 DR PIR; S09118; S09118.
 DR InterPro; IPR002895; Paramesium SA.
 DR Pfam; PF01508; Paramesium SA; 34.
 DR SMART; SM00639; PSA; 33.
 KW Signal; Repeat; Antigen; Membrane; GPI-anchor.
 FT SIGNAL 1 20 POTENTIAL
 FT CHAIN 21 2704 G SURFACE PROTEIN, ALLELIC FORM 168.
 FT DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.
 FT DOMAIN 1060 1424 VARIABLE DOMAIN, COMPRISES 4 ALMOST
 FT IDENTICAL REPEATS.
 SQ SEQUENCE 2704 AA; 278775 MW; 40EA0A0B18EE2119 CRC64;
 Query Match 11.5%; Score 665; DB 1; Length 2704;
 Best Local Similarity 23.9%; Pred. No. 1.5e-26;
 Matches 344; Conservative 43; Mismatches 487; Indels 568; Gaps 77;
 QY 3 GGCCCATCTGTTTTCAGCCAGTCG--CCAAGAA---TCA-----TGAAAG-TCGCC 47
 DB 258 GGCVTRTTCAAATTAQASCIKNSSGGDCYWTGTACVDKTCANAPTMTTNSACAGFTGCI 317
 QY 48 AGTGGCAGCACCCGCCCGCGCC-----GGGGGCC-----CAG-----CTGC 86
 DB 318 TKSG--GCVANGACSVANVQAACVKNSSNFCIDWTTCKEKTCAAPTNNTHDLCTSY 375
 QY 87 GCGCTGAAGCCCGGCAAGACA-----CGAGCGGTGC-----GGCGAGGT----- 127
 DB 376 LSTCTVKSG---GGCQNRSCANAPTMTTNDACEAYLTGNNCITKSGGCGVNTTCAAIT 432
 QY 128 -----GGTGC-----GC-----TGTCTGTCTGAGCAGA 150
 DB 433 LEACVKNSSGCTCFWDTASSCKDKTCVNPATNTHTDLCOAFLNTCTVNSTSAGCVK 492


```

RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RA "The complete genome sequence of Mycobacterium bovis."
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to a
CC frameshift in position 85.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z95890; CAB09322.1; -
CC EMBL; AB007040; -; NOT ANNOTATED CDS.
CC EMBL; BX248340; CAD94491.1; ALT_FRAME.
CC PIR; H70987; H70987.
CC TIGR; MT1807; -
CC Tuberculist; Rv1759c; -
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE_1.
CC Antigen; Repeat; Signal; Complete proteome.
KW SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 914 WAG22 ANTIGEN.
FT SEQUENCE 914 AA; 74354 MW; F6953CDBEB86AC8 CRC64;
Query Match 12.5%; Score 721.5; DB 1; Length 914;
Best Local Similarity 29.8%; Pred. No. 1e-29;
Matches 308; Conservative 20; Mismatches 377; Indels 329; Gaps 45;
QY 2 GGGCCCATCTCTTTC-----AGCCAGT-----CGCCAGATCATGAAGTCCCATGG 52
DB 83 GGGAYAAAEAAVTEPLINSINAPVLAATGRPLINGANGAP--GTGANGDAGWLIENG 140
QY 53 CAGCACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 111
DB 141 AGG-----SGAKGANGGAGGPGGAGG.FENG-GAGGAGGTATANNIGGAGGAGGS 190
QY 112 ----CGGTGCGGCG-----GAGTGTGTGCTCTGTCTGACGAGCGTGCCTATC 161
DB 191 AMLFGAGGAGGAGGAATSLVGGIGTGTGTG-GNAGML-----AGAAGAG-GAGGPFST 242
QY 162 TCGCGCTCGCG-----GGCGCGG-----GGCGCGCTGCTGCTGCTGCTGCTGCTG 208
DB 243 AGGAGGAGGAGGLFTTGGVGGAGGCGGTGGAGGAGGGLFGAGGAGGAGGFGDHTLGT 302
QY 209 GCAGCAGGTAAACGTCTCTCTACGATGAACGGCTGT--TACTACGCTCAAGGAG 266
DB 303 GGAGGDDGGGGLFGAGD-----GAGGGLTTGGAAGNGNAGTSLG-----AAGGAG 352
QY 267 CTGGTGCCACCTGCCCCAGAACCGCAGGTGACAGGTGAGATTCCTCAGCACGTC 326
DB 353 GTGG-----AGGTVFGGKGAGGAGG-----AGWLFSG 383
QY 327 ATCGACTACATCAGGACCTTCAGTTGGAGCTGAATCGGAATCCGAAGTTGGACCCCC 386
DB 384 G--GGGTGGFPAAGGQ-----GGVGSAGMLSGSGSGGAGGSGGAGTAAGGA----- 431
QY 387 GGGGCGCG-AGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 445
DB 432 GGAGGAPGPIGNGNGNGNGGSGGCTG-----GUGGAGGNAVLIENG 473
QY 446 -CCTGACGCGCGAGGCGGCGATG-----
DB 474 EGGIGALAGSGFGGFGGLLGLADGYNAPSTSPWHLQDILSPINEPTEALTRPLIG 533

```

```

QY 468 ----GTTCTCTGGGACGATCGCATCTTGTGTGCTGAAGCGCTCTCCCGCGGGA-----C 518
DB 534 NGDSGTPTGTGDDGAGGWLFGNGG-NGGAGAAGTNGSAG-----GAGGAGGILFG 582
QY 519 CGGCGGACCCCGACCCATCCAGGGGCG-----AAGAGGAATTACGTCTCTCTGG 567
DB 583 TGGAGG-----AGGVGTAGAGGAGGAGGSAFLIGSGGTGGVGAATTTGGVG-----GAGG 633
QY 568 -GTCTCCCCCAACGCGCTCG-CCGGATCTG-----AGGAGAACAAAGACCGAT--CG 616
DB 634 NAGLLIAGLGGCGGGAFTAGVTGGAGGTGGAGLFGANGGAGGAGGTGTAGGAGGAG 693
QY 617 GCGGCACTGCGCCCTTAACTGCATCCAGCTCGGGCTGAGGCTGAGGCACTGSC-GAGG 675
DB 694 GAGGLYAHGG-----TGGPGNGSGGTGAGGTGGAGG-----PGGLYAGG 733
QY 676 AGAGGCGCTCTCTCTCTCACACACTAGTACACAGAGACTTTAGGGGGTGGGATTCGA 735
DB 734 SGGAGHG-----GMAAGGGGVGNA----- 754
QY 736 CTCGTGTGTTCTTATTTTTTGAAGACAGACATTTTAAATAATGTCACGTTTGTGCTT 795
DB 755 -----GSILTNASGGAGGSGGSLSGKAGA-----GGAG-- 783
QY 796 CTCGATTTCTGAGGAATTTGTTTGTATTATATATATATATATATATATATATATATAT 855
DB 784 ----GSAGLFYGGGAGGNGGYSLNGTGGD-----GTGGAGQITGLRSGFGGAG 830
QY 856 TTGTTTACAATAG-----TTCGTGGGCTGTTTTTTTGTATTAAACAATAATTTAGA 911
DB 831 AGGASDTGAGNGGAGGAGGAGGAGGAGGAGGAG-----GDGATSGKGGAGNAVVIENG 883
QY 912 TGGTCAAAAAA 925
DB 884 NGGNAGKAGGTAGA 897
RESULT 4
LORI_MOUSE STANDARD; PRT; 481 AA.
AC P18185;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Loricrin.
GN LOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90275605; PubMed=2190691;
RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
RA Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinart P.M.,
RA Yuspa S.H., Roop D.R.;
RA "Identification of a major keratinocyte cell envelope protein,
RT loricrin."
RL Cell 61:1103-1112(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=BALE/c;
RX MEDLINE=95286248; PubMed=7738016;
RA Disipio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
RA Roop D.R.;
RA "The proximal promoter of the mouse loricrin gene contains a
RT functional AP-1 element and directs keratinocyte-specific but not
RT differentiation-specific expression."
RL J. Biol. Chem. 270:10792-10799(1995).
CC -!- FUNCTION: Major keratinocyte cell envelope protein.
CC -!- SUBUNIT: Monomers are crosslinked by disulfide and N-(gamma-
CC glutamyl) lysine isopeptide bonds.
CC -----

```



```

RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENEL; AL022022; CAAL1745.1;
CC PIR; F70806; F70806.
CC Tuberculin; RV3508;
CC InterPro; IPR000084; PE region.
CC Pfam; PF00934; PE; 1.
CC ProDom; PD001223; PE region; 1.
CC Hypothetical protein; Repeat; Signal; Complete proteome.
KW SIGNAL 1 30 POTENTIAL
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT CHAIN RV3508.
FT SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
SQ
Query Match 14.6%; Score 841; DB 1; Length 1901;
Best Local Similarity 32.3%; Pred. No. 2.1e-35;
Matches 335; Conservative 13; Mismatches 463; Indels 226; Gaps 40;
QY 1 GGGGCCCATCTGTTTCAGCCAGTCCCAAGATCATGAAGTCGCCAGTGGCAGCACCG 60
DB 180 GAGGWLFGVGGAGGVGGAG--GGTGGAGGPGGLTWGGGGAGGVGGAGGTGGAGGGA--E 235
QY 61 CCACCGCGCGCGCGCGCCAGCTGCGCTGAAGCGCGCAAGACAGCG--AGCGGTGC 118
DB 236 LFPAGAGAGAGTGG-----PGATG--GTGHHGVGGDGLWLPAGAGGAGCGGAGGAGS 290
QY 119 GGCC--GAGGTGGT---GGCGTGC---TGTCTGACAGAGCGTGCACATCTC--CGCT 168
DB 291 DGGALGTTGGTGGTGGAGGAGGAGGALLGAGGCGGGLGAGCGGCGTGGAGDGLVGGVGT 350
QY 169 GCGCG--GGCGCGGGCGCG--CTGCTGCTGCTGCGAGCGAGCGAGTAAAGTGC 225
DB 351 GKGGGVGGVAGLGAGGAGAGGOLFAGGAGAVGVGGTGGGAGGAGAGADAPA--STGL 409
QY 226 TGCTCTACGACATG-----AACGCTGTACTCACGCTCAAGAGAGTGGTCCCAACCT 280
DB 410 TGGTGFAGGAGGVGGCGNAIAGG--SGAGGTGGCGGAGGAGGWS 455
QY 281 CCCCCAGAACCGAAGGTGACAGGTGGAGATCTCCAGCAGCTCATGACTACTACG 340
DB 456 GADNAGSIGADG--GAGGTGGNAGAGGAGGAAGT-----GGTGGVGAAGKAGIGTGGQ 508
QY 341 GGA-----CCTTCAGTTGAGCTGAATCGGAATCCGAAGTTG-----GGACCCCGGG 389
DB 509 GGAGGAGSAGTATATGATGTTGGTGGGAGGAGGAGNTGVGTNGSGGCGT-----GGA 563
QY 390 GCGCGAGS-----GCTCCCGTCCGGCTCCGCTCAGCACCTCAACCGCGAGATC 440
DB 564 GGAGGAGGVADNPTGTTGGTGGGAGGAGGAGG-----GSSGAGGTN 608
QY 441 AGCCCTCTAGCGCGCGAGGCG-----GCATGCTTCTCTCGGAGCA--TCGCATCTGT 492
DB 609 GSGAGGTGGGAGGAGGAGGAGADNPTGIGAGGTG-----GTGAGAGAGAGATGTTGT 663
QY 493 -GTGCTGAAGCGCTCCCCAGGAGACCGCGGA--CCCCAGCCATCCAGGGGCAAGAGG 550
DB 664 GGAVGSGVNAIG-----GTGCTGGVGGAGGAGAGAGAGGAT-----GGAGFAGGAGG 712
QY 551 AATTACGTCTCTGTTGCTCTCCCCAACCGCGCTCCCGGATCTGAGGAGAACAGAC 610
DB 713 EGGAGGNSGVGTNGSGG-----AGGAGGKGTGGAGSGADNPTGAG 755

```

```

QY 611 CGATCGCGCGCCACTGCGCCCTTAACCTG--ATCCAGC-----CTGGGGCT----- 654
DB 756 FAGGAGGTGGAGAGGAGGAGATGCTGGTGVGATGATGAGTGGGCGGCGGAGGAGGAGG 815
QY 655 ---GAGGCTGAGGCACTGCG--GAGGAGAGGCGCTCTCTCTGACACACCTACTAGTCAC 709
DB 816 GFDGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 864
QY 710 CAGACACTTTTGGGGGTGGGATTCACATGCTGCTGTTTCTATTTTGTAAAGCAG-- 764
DB 865 VGGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 915
QY 765 ---ACATTTTAAATAATGTCAGTTCGTTGCTGCTCTCAGATTT----- 804
DB 916 GCGGAGAGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 975
QY 805 ---CTGAGGAAAT-----GCTTTGATTTGATATATTACATGATCACC 844
DB 976 DAGSGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1034
QY 845 GACTCAGATATGTTTTCATATAGTTCG-----TGGGGCTGTTTTTTTGTATTAAAC 899
DB 1035 GGINGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1094
QY 900 AAATAATTTAGATGCTG 916
DB 1095 AAGGCGGAGGAGGAGGAG 1111
RESULT 3
WA22_MYCTU
ID WA22_MYCTU STANDARD; PRT; 914 AA.
AC 006794;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WAG22 antigen precursor.
GN WAG22 OR WAG22B OR RV1759C OR MT1807 OR MTCY28.25C OR MB1789C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_Taxid=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Brown T., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;

```

RA Takagi T., Mizuno S.;
RT "Determination of the site of disulfide linkage between heavy and
light chains of silk fibroin produced by Bombyx mori";
RL Biochim. Biophys. Acta 1432:92-103(1999).
CC -!- FUNCTION: Forms the silk filament; a strong, inextensible,
insoluble and chemically inert fibre.
CC -!- SUBUNIT: Formed of two chains: heavy and light, that are linked by
a disulfide bond. Heavy-light chain assembly is essential for the
efficient intracellular transport and secretion of fibroin.
CC -!- TISSUE SPECIFICITY: Produced exclusively in the posterior (PSG)
section of silk glands.
CC -!- DOMAIN: Composed of antiparallel beta sheets. The strands of the
beta sheets run parallel to the fiber axis. Long stretches of silk
fibroin are composed of microcrystalline arrays of (-Gly-Ser-Gly-
Ala-Gly-Ala-)n interrupted by regions containing bulkier residues.
CC The fiber is composed of microcrystalline arrays alternating with
amorphous regions.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF226688; AAF76983.1; -;
CC EMBL: V00094; CAA23432.1; -;
CC EMBL: V00097; CAA23433.1; -;
CC EMBL: S74439; AAB31861.1; -;
CC EMBL: X13869; CAA32076.1; -;
CC EMBL: M35378; ABA27839.1; -;
CC EMBL: AB017362; BAA33147.1; -;
CC PIR: S01844; S01844.
CC Silk; Signal; Repeat.
CC SIGNAL 1 21
CC CHAIN 22 5263
CC FT CHAIN 22 5263
CC FT DOMAIN 149 5206
CC FT DISULFID 5244 5244
CC FT DISULFID 5260 5263
CC FT CONFLICT 10 10 C -> V (IN REF. 2).
CC SQ SEQUENCE 5263 AA; 391586 MW; 8EELD3A0A47440E CRC64;

Query Match 14.8%; Score 855.5; DB 1; Length 5263;
Best Local Similarity 31.5%; Pred. No. 8.5e-36;
Matches 322; Conservative 45; Mismatches 507; Indels 147; Gaps 32;

QY 1 GGGGCCCATCTCTTTGAGCCAGT---CGCCAGAAATCATGAAGT-CGCCAGTGGCAGC 56
DB 176 GAG- 234
QY 57 ACCGCCACCG 106
DB 235 AGYGAG 294
QY 107 AGCGAGCG-GTGGCGGCGAGGTGGTGGCGTGTCTGTCTGTGACGACGAGCGTGGCCATCTCGC 165
DB 295 AGSGAG- 346
QY 166 GCTCCG 224
DB 347 -----AG 400
QY 225 C-----TGCTCTACGATGAACGGCTGTGTTACTACCGCTTCAAGAGAGCTGTGCCCAACC 279
DB 401 AGAGYGAG- 454
QY 280 TGCCCCAGAACCGAGTGAGCAAGTGAGGAGATTCTCCAGCAGCTCATCGACTACATCA 339
DB 455 AG 507
QY 340 GGGACCTTCAGTTGAGCTGAATCCGAATCCGAATCCGAATCCGAATCCGAATCCGAATCCGA 399

DB 508 GAG- 566
QY 400 TGCCGGTCCGGCGCTCCGCTCAGCACCTTCAACGCGGAGA---TCAGCGCCCTGACGCGCG 456
DB 567 SGASGAG 626
QY 457 AG-----GGG-CCATCGCTTCTCGGACGATCGCATC----- 488
DB 627 AGVGAG- 686
QY 489 -----TTGTGTCGCTGA-----AGCGCCTCCCGCGGACCG----- 520
DB 687 YSRDGYEYAWSSFFGTGSGAG 746
QY 521 -----CGGACCCCGAGCCATCCAGGGGCGGCAAGAGGAATTACGTCTCTGTGGGTCT 571
DB 747 VGYGAG- 803
QY 572 CCCCCAACCGCGCTCCGCGGATCTGAGGGAGAACAGACCGATCGCGCGCACTGCGCCCC 631
DB 804 GAG 863
QY 632 TTAATCTGATCCAGCCTGGGCTGAGCTGAG---GCACCTGGCGAG-GAGAGAGCGCTCC 687
DB 864 GSGAASG-----AGAGSGAG 918
QY 688 TCTCTGCACACTACTAGTCA-CCAGAGACTTTAGGGGTGGGATTCCTCTGTTGTTT 746
DB 919 AGAGSGAG 973
QY 747 CTATTTTGTAAAGACGACACATTTTAAATAATGTCTACGTTTG-GTGCTTCTCAGATTTC 805
DB 974 GAGSGAG- 1029
QY 806 TGAG-GAAATGCTTTGTATTGTATTATATATATATATATATATATATATATATATATAT 864
DB 1030 -GYGAG 1087
QY 865 AATAGTCTCTGGGCGCTGTTTTTTTATTATAAATAATATATATATATATATATATATAT 924
DB 1088 GAGSG---AGAGSGAG 1144
QY 925 A 925
DB 1145 S 1145

RESULT 2
Y208_MYCTU
ID Y208_MYCTU STANDARD; PRT; 1901 AA.
AC O53553;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv3508 precursor.
GN Rv3508 OR MY023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churche C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitthead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:11:40 ; Search time 17 Seconds
(without alignments)
2836.292 Million cell updates/sec

Title: X77956
Perfect score: 5766
Sequence: 1 GGGGCCCATCTGTTTCAGC.....TTAGATGCGAAAAA 926

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	855.5	14.8	5263	1 FBOH_BOMMO	P05790 bombyx mori
2	841	14.6	1901	1 YZ08_MYCTU	O53553 mycobacteri
3	721.5	12.5	914	1 WAZ2_MYCTU	O06794 mycobacteri
4	706	12.2	481	1 LORI_MOUSE	P18165 mus musculu
5	680.5	11.8	2715	1 G156_PAPPR	P13837 paramesium
6	665	11.5	2704	1 G168_PAPPR	P17053 paramesium
7	647.5	11.2	957	1 Y278_MYCTU	P56877 mycobacteri
8	635.5	11.0	747	1 SPDI_NEPCL	P19837 nephila cla
9	629	10.9	778	1 YQ34_MYCTU	P71333 mycobacteri
10	627	10.9	801	1 Y747_MYCTU	O53810 mycobacteri
11	618.5	10.7	672	1 PHX5_MOUSE	P08399 mus musculu
12	570.5	9.9	2321	1 NTC3_HUMAN	Q9um47 homo sapien
13	562.5	9.8	2319	1 NTC3_RAT	Q9r172 rattus norv
14	554	9.6	2531	1 NTC1_MOUSE	Q07008 rattus norv
15	553	9.6	2531	1 NTC1_MOUSE	Q01705 mus musculu
16	551.5	9.6	2318	1 NTC3_MOUSE	Q61982 mus musculu
17	547.5	9.5	4289	1 TENX_HUMAN	P22105 homo sapien
18	544.5	9.4	2556	1 NTC1_HUMAN	P46531 homo sapien
19	543	9.4	1700	1 BAR3_CHITE	Q03376 chironomus
20	527.5	9.1	641	1 EBNI_EBV	P03211 Epstein-bar
21	518	9.0	2003	1 NTC4_HUMAN	Q99466 homo sapien
22	517.5	9.0	1964	1 NTC4_MOUSE	P31695 mus musculu
23	510	8.8	1046	1 PSTA_DICDI	P11976 dictyosteli
24	509	8.8	1064	1 FBPI_STRPU	P10079 strongyloce
25	502.5	8.7	2437	1 NTC1_BRARE	P46530 brachydanio
26	501.5	8.7	5376	1 ZAN_MOUSE	O08799 mus musculu
27	496	8.6	2524	1 NTCX_XENLA	P21783 xenopus lae
28	493.5	8.5	465	1 GRP2_PHAVU	P10496 phaseolus v
29	482.5	8.5	2471	1 NTC2_RAT	Q9qk30 rattus norv
30	486.5	8.4	603	1 YD25_MYCTU	Q10637 mycobacteri
31	484.5	8.4	2471	1 NTC2_HUMAN	Q04721 homo sapien
32	481	8.3	384	1 GRP1_PETHY	P09789 petunia hyb
33	478	8.3	2470	1 NTC2_MOUSE	O35516 mus musculu

34 465 8.1 2703 1 NOTC_DROME
35 452 7.8 860 1 ELS_MOUSE
36 450.5 7.8 349 1 GRP_ARATH
37 445 7.7 1150 1 APMT_PIG
38 431 7.5 316 1 LORI_HUMAN
39 426 7.4 182 1 KRUC_SHEEP
40 422.5 7.3 194 1 KRUB_HUMAN
41 418 7.2 498 1 Y118_MYCTU
42 416.5 7.2 713 1 TSA4_GIALA
43 414 7.2 864 1 ELS_RAT
44 407 7.1 178 1 CHHC_BOMMO
45 403.5 7.0 2907 1 FBN2_MOUSE

ALIGNMENTS

RESULT 1
FBOH_BOMMO STANDARD; PRT; 5263 AA.
AC P05790; Q17220; Q26379;
DT 01-NOV-1988 (Rel. 09, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).
GN FIBH.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330362; PubMed=10871375;
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Enault C.,
RT Yang T., Jacquet M., Janin J., Duguet M., Ferraso R., Li Z.-G.;
RL "Fine organization of Bombyx mori fibroin heavy chain gene.";
RN [2]
RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT flanking, mRNA coding, entire intervening and fibroin protein coding
RT regions.";
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
RT "Structural analysis of the fibroin gene at the 5' end and its
RT surrounding regions.";
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=89094868; PubMed=3210244;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RT structure of silk fibroin mRNA.";
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RT gene.";
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RC STRAIN=J-139;
RX MEDLINE=99296390; PubMed=10366732;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,

P07207 drosophila
P54320 mus musculu
P27483 arabidopsis
P12021 sus scrofa
P23490 homo sapien
P26372 ovis aries
O75690 homo sapien
Q50615 mycobacteri
Q99372 rattus norv
P20730 bombyx mori
Q61555 mus musculu

```

Db 901 YRAGCTNTVGAASSASCTLDCTLKT-----GSLTFADQCALDSTCSVKKDGTCIVIQST 956
QY 457 AGGCGGCAATGCGTTCCTCCGACGACATCG-----ATCTTGTGTGG-----CT 498
Db 957 CAGYGSTATNCFRSSASCTAGYCAMNTNCSQSVTGAECAPVTLGLTGLDHSKCOLYHSSCT 1016
QY 499 GAA-CCGC---CTCCCCCAGGAGCCGCGGAC-----CC-----CA-----GCCATCCAG 539
Db 1017 SLKDTGCGQYKTAGSSVATGNTCANSVQKCPDADTCLRFANCASITGTGLTNTICVT 1076
QY 540 GGGGCAAGAGAAATTAGC-----TGCTCTGTGGTC-----TCCC-----C 575
Db 1077 YDPGCVANVG---TACQEKLATCAAYLTQNSCTSTAGTCAMSGSACLTVVDANVATEC 1133
QY 576 CAAACGCGCTCGCGG-GATCTGAGGGAG-----AACAGACCGATCG-GCGGCCACTGCG 628
Db 1134 AVITGTGLTNAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCSQSDAGLCAMSSGA 1193
QY 629 C-----CTTAACTGC-----ATCC 643
Db 1194 CLTVVDANVATECPYITGTGLTNAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCS 1253
QY 644 ---AGCCT-GGGGCTGAGGCTGAGGCACTGGCGAGGAGAGGGGCGCTCTCTCTGCGAC--- 696
Db 1254 QSDAGLCAMSGSACLTVVDANVATECPYITGTGLTDAICAG--YNAKCTVNRAGTACQK 1311
QY 697 -----ACCTACTAGTACACAGAGACTTTAGGGGTGGGATTCACACTCGTGTGTT 745
Db 1312 EALCATYAAVQATCSQSDAGLCAM--SGSACLTV-----DANVATECPYITGTGLT 1361
QY 746 T-----CTATTT-----TTGAAAAGC--AGACA-----T 768
Db 1362 NAICAGYNAKCTVNEAGTACQKKEALCATYAAVQATCSQSDAGLCAMSGSACLTVVDANV 1421
QY 769 TTTAAAAAATGTT---CA-----CGTTTGGTC---TTC-----TCAGATT----- 803
Db 1422 ATECAVITGTGLTDAICAGYNAKCTNLKDGTCQCDERATCKLYTTONKCTSQTTGPLSCL 1481
QY 804 -----TCTG-----AGGAAATTGCTT-----TGTATTG 826
Db 1482 WFDNSCSPITDVTCSAIVQSLDHAQCOAYSTGCTSVSDGSKCODFKTTCQYAGTALSC 1541
QY 827 TATATTACAAATGATC-ACCGACTGAGAATATTGTTT-----TACA----- 865
Db 1542 TKTATSKCYLQNSNCITISNVATDCAKITGSAGTITVEICQSVNTGCSVNRARSACVQQQ 1601
QY 866 -----ATAGTTCTGTGGGCTGTTTTTTTGTATTAAACAAATAATTAGATGGTCAAA 919
Db 1602 AOCSGVTSAMTSCYKSGAGLCIASINTDT-----ACVAATAATTCDAVVLGTGNY 1652
QY 920 AA 921
Db 1653 SA 1654

```

Search completed: May 7, 2004, 15:20:01
Job time: 39 secs

C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.9%; Score 686.5; DB 2; Length 1660;
Best Local Similarity 29.8%; Pred. No. 1e-24;
Matches 305; Conservative 20; Mismatches 460; Indels 239; Gaps 39;

QY 1 GGGGCCCCATTTCTTTTTCAGCCAGTCGCCAAGAA--TCATGAAGTCGCCAGTGGCAGCAC 58
DB 141 GNG--AGAVGVGAGAGAGLFIIGGAGGAGAGAPGGTGTGWLGGGGVGGMG- 196
QY 59 CGCCACCGCGCGCGGCCAGCTGCGCGTGAAGCCGCGGAGAGACAGACGAGCGGTGC 118
DB 197 -----GAGG-----GAGGAGGAGNAGLFNGGAGGAGGAGGAGGAGG 232
QY 119 GGCC----GAGTGGTG-CGCTGTCTGTGACAGAGCGTGGCCATCTCGCGTGGCGG 173
DB 233 NAGFHGGGAGGAGGAGAGAGATFGQDGAAGVAGSDGAG-----GGLAGSDG 284
QY 174 -GGCGCGCGCGCGCGCTGCTCCCTGCTGTGACGAGCAGAGGTAAAGTGTGCTG-CTCT 231
DB 285 DGGAGGVGGNG-SRG--GWLNGGAGGVGGVGGAGGAGAGGAGGATGINGPAGIS 340
QY 232 ACACATGAACGCTGTACTCAGCTCAGGCTCAGGAGCTGGTCCACCCCTGCCCGACACC 291
DB 341 AAGGDDGAGGAGGAGNG-----GVGAGGAGGS-----AGLGYVGRAGD 381
QY 292 GCAAGGTGACAA--GGTGGAGATTTCTCCAGCAGCTCATGCACTACATCAGGACCTCA 349
DB 382 GGAGGGGGLGAPDGGAGGAGGSLAAGDGGAGGAGGDPGLGA-----GGA-----G 430
QY 350 GTTGGAGCTGAA--CTCGGAATCGAAGTTGG--GACCCCGCGGGCGGAGGCTGCGGT 406
DB 431 GASGGAGARAGANGLAAGNDPVSNGGNGGAGAHAPVAGHGGNGAGGN-----GGL 485
QY 407 CCGGGCTCCGCTCAGCACCCCTCAACGCGGAGATCAG-----CGCCCT-----GACGCG 454
DB 486 VDDG-----AGHGGDGAAGAGVADMTAIFLGSSGTPEGDDGNGGA 527
QY 455 CGAGGCGCATGCGTTCCTCGGAGCATCG-----CATTTGTGTGCTGAAGCCCTC 508
DB 528 GGAGGAGAGAGDGGAGGAGGAGGAGGAGHFNVLVSDGNGDGGAGRG-----583
QY 509 CCCAGGAGCGCGGAGCCCGACCATCCAGGGGGAAGAGGAATTACGTGCTCTGTGG- 567
DB 584 -----GDGA--GGAGD--APAGAGGQGVGGDGG--AGGAGAP-----GNG-----GSGR 626
QY 568 GTCTCCCCCAACGCGCTCGCGCATCTGAGGAGAGAAAGAGCCGAT-----CGCGGGCC 622
DB 627 GD-----NAFKDGDGAGDGDGDPGAGGKGGAGGAGATGEGTGTATGATVHSGNGG-- 677
QY 623 ACTGGCCCTTAACTGCATCCAGCTCGGCTGAGGCTGAGGCAC-----667
DB 678 -KGNAGADATVAGANG-----GKGAGNGGLVGDGGAGDGGGAGAGANGVGEDGADG 732
QY 668 -----TGCGAGGAGAGGCGCTCTCTCTGTCACACCTACTAGTCTAC 709
DB 733 TLSGPBGSEANGGQGGVGGGAGGAGDGGAGSSALGSGGNGRGDAGQAGGAGG 792
QY 710 CAGAGACTTTAGGGGTGGGATTCATCTGCTGTTCTTCTATTCTTCAAAA-----760
DB 793 AGGAGSVSGDGPFGGAGGAGAGAGAGAGGAGGAGSAGSADSAFVAGGAGKGGDGVGG 852
QY 761 -----GCAGACATTTTAAA-----AAATGCTCAGCTTTGGTCTCTCAGATTTC 805
DB 853 VGGDGGPGDGGAGAGAPAGQVGSVGVDGGLGAGGNGDGGHGDGGDGGDGD 912
QY 806 TGAGAAATTTGTTTATTGTATATATACATATGATCAGGACTCAGAAATTTGTTTACA 855
DB 913 FGAGGLGLGDSNGNTR-----AASGVDSADHGPSSGNGGNGGNGAGQASVAG 961
QY 856 ATAGTTCTG-----TGGGGCTGTTTTTTTGTATTATAACAAATAATTAGATGCTGAAA 920
DB 962 GAGNGGDDGNAGRVGDGAGGNGDGAAGANGANGSAGPSDALALGQPGNGGQGDAGQ 1021

QY 921 AAAA 924
DB 1022 AGGA 1025

RESULT 15

S09118
G surface protein 168 - Paramesidium primaurelia
C:Species: Paramesidium primaurelia
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 07-Dec-1999
C:Accession: S09118
R:Prat, A.
J. Mol. Biol. 211, 521-535, 1990
A:Title: Conserved sequences flank variable tandem repeats in two alleles of the G sur
A:Reference number: S09118; MUID:90172419; PMID:2308165
A:Accession: S09118
A:Molecule type: DNA
A:Residues: 1-2704 <PRA>
A:Cross-references: EMBL:X52133; NID:g10049; PIDN:CAA36378.1; PID:g578473
C:Genetics:
A:Genetic code: SGC5
C:Superfamily: G surface protein

Query Match 11.5%; Score 665; DB 2; Length 2704;
Best Local Similarity 23.9%; Pred. No. 1.3e-23;
Matches 344; Conservative 43; Mismatches 487; Indels 568; Gaps 77;

QY 3 GGGCCATCTCTTTTCAAGCCAGTCG--CCAAGAA--TCA-----TGAAG-TGCCC 47
DB 258 GGCVTTRTTCAAAATQAACIKNSGGDCVWTGTACVDTKTCANAPTMTTNSACAGFTGCI 317
QY 48 AGTGGCAGCCGCCACCGCGCC--GCGGGCC--CAG-----CTGC 86
DB 318 TKS--GCVANGACSVANVOAACVKNSNFDCTWDTCKEKTCAANAPTNNTHDLCTSY 375
QY 87 GCGCTGAAGCGCCAGACACA-----GCAGCGGTGC-----GGCGAGGT- 127
DB 376 LSTCTVKSG--GGQNRSCANAPTMTTNDACEAYLTGNNCTKSGGCVTNTTCAIT 432
QY 128 -----GCTGC-----GC-----TGCTGCTCAGCAGA 150
DB 433 LEAACVKNSGSTCFWDTASSCKDKTCVNAPNTNTHDLCOAFINTCTVNSTAGCVEX 492
QY 151 GCGTGGCCATCTCGCGCTGCGGGGC--GCGGGGCGCGCTGCTGCTGCC--TGCTG 203
DB 493 TCENSLVLAICDKDTSSRACIWKGYKKQCVLAS--SATTTHADQTYHSTCTL 545
QY 204 GAGGAGCA-----GCGGT-AAACGTGCTCTACGACATGAAGCG--TCTTACTCAG 256
DB 546 SNSGTGCVPLPKCEAITIEAACNLKANGQPCGWNGSQCIDKACSTASKTFTTTSQCTGH 605
QY 257 CCTCAAGGAGCTGGT-----GC-----CCACCTGCC-----CCAGAA 289
DB 606 ISTCVANNPVTNGLTIQGDLPSTCAARKSENCEIARVGPPTCLWVSSSTSCVBKS 665
QY 290 CGCAAGGTGAGCAAGTGGAGATTC--TCCA-----GC-----A 322
DB 666 CATASTVGTGTALSAGGTFSGCQTYLNTCISNNTADGCIAPKSSCSLSVSNCRDGSKA 725
QY 323 CCTCATCGACTACATCA-----GGGAC-----CTT-----CA 349
DB 726 SDCFTWNGSSCVDKTCANITLTSHASCYSIFNQCTVWNGGTACCTLATACTSYTQENCK 785
QY 350 GTTGGAGC--TGAATCCGAATCCGAAGTGGAGCCCG-----GGGGCC 393
DB 786 FRTSTNKNKCVMTGLACR--NATCADAPDPTTAYDSDECLAYPTPSETCTTVVYKVAQGC 842
QY 394 GAGGCG-----TGCGGCTCCGGGTCC--415
DB 843 SKSANCSYMTSAQCHKITLNTANDCKWIVDRCYALSSFATGAC--TTFKGNKTMCEG 900
QY 416 ---GCT-----CAGCACCCCTCAACGCGGAGATCAG-----GCCTGACG--GC-----CG 456

A; Gene: Rv2490c

A;Accession: B70812
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-882 <COI>
A;Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17640.1; PID:g2916
A;Experimental source: strain H37RV
C;Genetics:
C;Gene: RV0834c
C;Superfamily: elastin

Query Match 12.1%; Score 699; DB 2; Length 882;
Best Local Similarity 30.8%; Pred. No. 2e-25;
Matches 294; Conservative 15; Mismatches 369; Indels 2

37 TGAAGTGGCCAGTGGCAGCACCG-CCACGCCGCCGCCGCCGCCAGCTGGCGCTGAAG 95
132 TGEAGGCGWLGNCGNGCGSGAPGCTGGAGGAAGLLGHCG-----TG-GAGGTGAAG 182

y
 96 GCGCGCAAGACACGACGAGCGGTGCGGGCGAGGTGGT-----GGCTGTCTGTCTTGAGCACAG 151
 183 G--KGGTGGWLWGGG-GAGGAG-GSCGGGCGAGGNALMFTGGNGGAGGAASGVNGGVG 238
 b

a

CGTGGCCATCTCGGCTGC GGGGCGCGCGGGCGCGCTGCTGCCCTGCTGG - ACAGAG 208
152

b

GAGGAGCAATVATGGAGCAGGAATTGTGAGGAGGAGN - ALGHFLGLGGGGGGSDSAMSG 296
239 - GAGGAGGAGALVATGGAGCAGGAATTGTGAGGAGGAGN - ALGHFLGLGGGGGGSDSAMSG 296

210 CAGCAGTAAAC-----GTGCTGCTCAGACATGAA CGCGTGTTACTCAGCGCTCAA 262

297 GAGGAGCGSGAASPFGIDIGGAG-----CHGGGATNGGAG-----A 335

263 GGAGCTGGTGCCCAACCCCTGCCCCCGAACCCTGAAGGTGCAGATTTCTCCAGCA 322
|||||
266 CTTCGCCCTTCAATDINISWCG-----GGNG-GAATTGTGCAGCGGFAVAPDF 381

323 CGTCATCGACTACATCAGGAGACCTTCAGTTGGAGCTGAACTCGGAATCCGAAAGTTGGAC 382

383 CCCCCGGGGCCGAGGGGCTCCCGGTCGGGGTCGGCTCAGCACCCCTCAACGGCGAGATCAG 442

443 CGCCCTGACGGCGAGCGGCATGCGTTCCTCGGAGCATGCCATCTTGTCCTCTGAAG 502

484 TG-----GGGCGGAATDQ-----GNGCAGC 50-2257-1096
503 CGCTCCCCAGGCAC--CGCGCGACCCCAGCCATCCAGGGGCAAGAGGAATTACGTGC 560

519 -----GHHGGGASVGTGGSSG-----AGGDDGGFVGGAGGNGGMA-GIGV 519
561 TCTGTGGGT-CTCCCCCRAAGCGGCGCTCGCGGATCTGAG--GGAGAAC-----AAGA 609

556 GVNCGANGNGGSATGALAAVGG---GAGAAAGGDATCTGGFGGAGGSGARGLIFALGGAGA 611
bb
610 CGGATCGGCGGCGCACTGGCCCTTAACTGCATCCAGC-CTGGGGCTGA----- 656
by

612 AGGDASTGVGGPGGPGGTG---TASSPFGIAIAICGAGAQQGAGTSGATGGAGGADVFE 668

657 -----GGCTGAGGCA-----CTGCCGA-CGAGAG-----GSGCCTCCTCTCT 692

669 IAVLGLFGGAAGAGGAAITGDGATCGAGGCGGAGAGATNFLGFSVLHGAG-----719

693 GCACACCTACTAGTACACGAGAGACTTTAGGGGGTGGGATTCCACTCGTGTGTTCTATT752

720 -----CAGGTATGTGNGGAGGGGLSSPVILGICIG----- 751

752 -----CAGGTATGTGNGGAGGGGLSSPVILGICIG----- 783

784 -----CAGGTATGTGNGGAGGGGLSSPVILGICIG----- 815

752 ---GAGDGGALGVLGGMGDDGGGEAAVAVGIAVGGAG-----GAGGA 793

794 APTCNGGAGCG--GDALGLVGVGNGG-----NAGTCFGANTCGNGGTTIVVNGMLAP 846

361 LGLDGNAPASTNPLHTAQOQALAAVNAPIQAVTGRPLINGANGAPGSGPAGGGWLF 420
 441 AGCGCCCTGACGCCGAGCGGCGATGCGTCTCTGCGGACGATCGCATCTTGTGCGCTGA 500
 421 GGGGTGSGVSGAGGCGGAGGILFGAG-----GAGGAGGA-----VTGTGATGSGGG 468
 501 AGCGCTCCCGCAGGACCGGCGGACCCAGCCATCCAGGGGCGCAAGAGGAATACGTGC 560
 469 AGGG-----ALLFGAGA-----GAGGGS-----SGIGFAAGGAGPG-----GAGG 506
 561 TCTGTGGTCTCTCCCGACCGCTCGCGGATCTGAGGAGAAACAACCGATCGCGG 620
 507 LFNCGGAG-----GAGGSGVSGAGGCGGAGGAGGILFAGGAGGAGG 548
 621 CCACTGGCCCTTAATGTCATCCAGCTGGGCTGAGGCTGAGGCACTGGCG--AGGAGA 678
 549 GGNVNG-----GAGGAGVGGGLFAGGAGGSGGSGVAGDSGA 586
 679 GGGCGTCTCTCTGACACCTACTAGTACACGAGACTTTAGGGGGTGGATCCATCTC 738
 587 GGNAG-----LLAPLAGAGGGGCGGDFDTGGAGGPGDA-----GLLV 625
 739 GTGTGTTCTATTTTGAAGACACACATTTTAAATATGTCAGTTTGTGCTTCTC 798
 626 GSG-----GVGAGGPGFLTGGPG-----AA 646
 799 AGATTTCTGAGAAATGCTTTGTTATTTATGATATTAACAATGATCAGCAGTGAATATG 858
 647 GSDAGLLFGSGAGGAGGSGRTDLGGAGGAGGKAGLIGNGN-----GGAGGAGNGG 699
 859 TTTTACAAATAGTCTGTGGGCTGTTTGTGTTTATTAACAATAATTTAGATGTT 915
 700 GDGPGGAATFLGNGGNGNGGTGT-----SAGSPGAGAGGS 737

RESULT 10
 A70934
 hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70934
 R: Cole, S. T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrall, B. G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: A70934
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-1306 <COL>
 A: Cross-references: GB: AL021942; GB: AL123456; NID: G3242298; PIDN: CAA17449.1; PID: G290963
 A: Experimental source: strain H37Rv
 C: Genetics:
 A: Gene: Rv0578c
 C: Superfamily: collagen alpha 1(IV) chain

Query Match 12.3%; Score 712; DB 2; Length 1306;
 Best Local Similarity 30.2%; Pred. No. 6.4e-26;
 Matches 332; Conservative 20; Mismatches 464; Indels 282; Gaps 39;

QY 1 GGGGCCCATCTGTTTCAGCCAGTCGCCAAGAATCATGAAA--GTCGCCAGTGGCAGCAC 58
 DB 167 GAGG-----VGITGGAG-----GHGAGGWLNGGAGGFGGAGVGVNGGAGGTA- 212
 QY 59 GCGCACCCCGCGCGG--GCCCCAGTGC-----GGCTGAAGCGCGCAACACACGA 111
 DB 213 -----GLFGVGGAGGAGNGIAGVTGTSATSTPGSGTAGGAGGICGNGGAGGAGGVLM 265
 QY 112 GCGGTGCGGCGAGGTGCT--GCGTGTCTGTCTGAGCAGACGCTGGCCATCTCG----- 164
 DB 266 GNGGN--GGAGGEGPFGAGAGAGAGAHATNLGADGGGNGGNGGAGGTGGVGGPGGH 323

QY 165 --CGCTGCCGGGGCGCGG-----GGCGCGCTCTGCCCTGCCCTGCGACGAGCAGCAG 215
 DB 324 GLLGLGSHGAGGAGGSGGDPGNGATGCTWGNLGAAGTGGNGNPFAGGAGGAG 383
 QY 216 GTAAACGTGTCTCTACACATGAACGCTGTACTACGCTCCTCAAGAGAGCTGTGCCCC 275
 DB 384 G--ASVGSAGHAN--GAPGTTSTSGNGSGDGKADAISSG--QTGANGRGGGSGGQ--- 435
 QY 276 ACCTGCCCCAGAACCGCAAGGTGAGC-----AAGGTGGAGATTCTCCAGCACCTCAT 328
 DB 436 -----VNGGAGGAGRG--GAGGLGFSEAPRPGAGGTGGAGN-----GGTQAGDGGT 485
 QY 329 CGACTACATCAGGACCTTCAGTTGAGCTGAATCCGAATCCGAATTTGGACCCCGG 388
 DB 486 GGA-----GGA-----GGDGSAGAGSIFGNASAPGAAGSPGNGNGGPGG 527
 QY 389 GGGCGGAG-----CCTCCGCTCGGCTCCGCT-----CAGCA 423
 DB 528 AGGEGGAGLALAAASQNGSQGAGGAGGNGGTGNGHGAAGALVNGVGGAGHG 587
 QY 424 CCCTCAACGCGAGATCAGCGCC--CTGACGCGGAGGCGGCGATCGTTCTCT----- 473
 DB 588 GPGVGGAGGCGGSGSTFGANGAFNPTTSGNGNGGAGADATGFGQTGASGGRGDGG 647
 QY 474 --GGGACGATCGCATCTTGTCTGCTGAAGCGCTCCCCAGGACCGCGGACCCCG 531
 DB 648 LVNGGAGGAGGNGSKGLPLGLRLGNPLDG-----GTGNGGAGGSGGA---WAG 695
 QY 532 CCATCCAGGGCGCAAGAGGAATTAACGTCTCTGTGGCTCTCCCCCAACGCGCTCGCGG 591
 DB 696 NGTGTGAGTGGVGTGSGSDGVNGSSAGADHPGCT-----GGVGG 738
 QY 592 ATCTGAGGAGAACAGACCGATCGCGCGGCACTGC-----GCCCTTAAGTCATC 642
 DB 739 TGGGCGDGGAAAPNGVAGSQGPGAGDGTGGVNGNGRGIDGADGATAGARG--QDG 797
 QY 643 CAGCTCGGGCTGAGGCTGAGGCACTGGCGAGGAGAGGCG-----GCTCCTCTCTG 693
 DB 798 GAGGAGGKRGCTGGTGGGAGPAGTTGSGAGGNGSGGTGGDPGCGANGANGSVFTNG 857
 QY 694 CACACTACTAGTCAACAGAGACTT-----TAGGGCTGGATTCCACTCGT-----GTGTT 745
 DB 858 ICGNGGNGNAGPSGAGSGGAGSTFGATGSSSIHVNNGNGNGGNGDHLSNGAAG 917
 QY 719 -----TAGGGCTGGATTCCACTCGT-----GTGTT 745
 DB 918 NGNGGNGSLRSGGAGGNGGNGNARGMGDDGTGGAGGAGGAGGAGGAGG 977
 QY 746 TCTATTTTGAAGAGCAGACATTTTAA-----AAATGTCACGTTTGTGCTT 795
 DB 978 GSDGNPGAITGSGGCGGSDGVGGGAGVAGDAGDGRGAGGTGCTGLRG--TTGATGATG 1036
 QY 796 CTCAGATTCTGAGGAATTCCTTTGTTATTTATTTATTTATTTATTTATTTATTTATTT 851
 DB 1037 TFDAGA-----DOHNGNGTGVGGTGGAGGNGGAGGAGGAGGAGGAGGAGGAGGAG 1092
 QY 852 AATATTGT-----TTTACAATAGTTCTCTGGGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 903
 DB 1093 GAGGTGTTGGDGRGAGHGTGTFSSLAGT--GGTGGNGGTG-----GTGTGGAGGAGGT 1143
 QY 904 AATTAGATGCTGAAAA 921
 DB 1144 G--STLGATGATGAAGA 1159

RESULT 11

A35628
 loricrin - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 21-Sep-1990 #sequence revision 21-Sep-1990 #text_change 13-Aug-1999
 C:Accession: A35628
 R: Mehrel, T.; Hohl, D.; Rothnagel, J. A.; Longley, M. A.; Burdman, D.; Cheng, C.; Lichti,

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70987
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-914 <COL>
A;Cross-references: GB:295890; GB:AL123456; NID:93242245; PIDN:CA809322.1; PID:92131027
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv1759c; wag22
C;Superfamily: elastin

Query Match 12.5%; Score 721.5; DB 2; Length 914;
Best Local Similarity 29.8%; Pred. No. 1.9e-26;
Matches 308; Conservative 20; Mismatches 377; Indels 329; Gaps 45;

QY 2 GGGCCCATCTGTTTC-----AGCCAGT-----GCCCAAGAAATCATGAAAGTCCGCCAGTGG 52
DB 83 GGGAYAAEAATVPLNLSINAPVLAATGRPLINGANGAP--GTGANGDAGWLINGG 140
QY 53 CAGCACCGCCACCGCCCGCCCGCCAGCTCGCGCTGAAGCCGCGCAAGCACGGA- 111
DB 141 AGG-----SGAKGANGAGGPGGAAGLFGNG--GAGGAGTATANNIGGAGGAGS 190
QY 112 ---GCGTGTGGGCG-----GAGGTGTGCGCTGTCTGTCTCAGCAGAGCGTGGCCATC 161
DB 191 AMLFAGAGAGAGGAATSLVGGTGTGTG--GNAAGL-----AGAAGAG--GAGGFPST 242
QY 162 TCGCGCTGCCGG-----GGCGCGG-----GGCGCGCTCGCTGCCCTGCTG-----GACGA 208
DB 243 AGGAGGAGGAGGLFTTGGVGGAGGCGGHTGGAGGAGGAGGLFGAGMGAGGFGDHTLGT 302
QY 209 GCAGCAGTAACCTGTCTCTACGACATGAACGGCTGT--TACTCAGCCTCAAGGAG 266
DB 303 GGAGGCGGGGLFGAGDG-----GAGGSLTTGAGNAGNAGTSLG-----AAGGAG 352
QY 267 CTGCTGCCACCTGCCCGGCTCGCGCTCCGCTCAGCACCTCAAGCGGAGATCAGCGC 445
DB 353 GTGG-----AGGTGFGGKGGAGGAGGN-----AGMLFGSG 383
QY 327 ATCGACTACATCAGGACCTTCAGTTGGAGCTGACTCGAATCCGAACTGGAGCCCCC 386
DB 384 G--GGGTGGTGAAGG-----GGVGSAGMLSGSGSGGAGGSGGPGTAAGGA----- 431
QY 387 GGGGGCGG--AGGGTGTGCGGTCCGGCTCCGCTCAGCACCTCAAGCGGAGATCAGCGC 445
DB 432 GGAGGAPFLINGNGNGGSGESGTG-----GVGGAGNAVLINGG 473
QY 446 -CTGACGCGCGAGCGCGCATG----- 467
DB 474 EGGIGALAGSGFGGFGGLLADGVNAPESTSPHNLQDILSFINEPTEALTGRPLIG 533
QY 468 ---GTTCTCTCGGACGATCGCATCTTGTGTGCTGAAGCGCTCCCGCAGGGA-----C 518
DB 534 NGDSGTGTGDDGGAGGWLFGNG--NGGAGAAGTNGSAG-----GAGGAGGILFG 582
QY 519 CGGCGGACCCAGCCATCCAGGGGCG-----AAGAGAAATTAAGTCTCTGTGG 567
DB 583 TGGAGG-----AGGVGTAGAGGAGGAGGSAFLIGSGGTGVGGAATTTGGVG-----GAGG 633
QY 568 --GTCTCCCAAGCCGCTCG--CCGGATCTG-----AGGGAACAAGACCGCAT--CG 616
DB 634 NAGLLIAGAGLGGCGGAGTGTAGTTGGAGTGTGAGLFGANGAGGAGGTGTAGAGGAG 693
QY 617 CGCGCCACTGCGCCCTTAATGTGATCCAGCTGCGGTGAGGTGAGGCACTGCG--GAGG 675
DB 694 GAGGLYAHGG-----TGPGGNGGSGTGAAGTGGAGG---PGGLYGAGG 733
QY 676 AGAGGCGGCTCTCTGTGCACACCTACTAGTACCAGAGACTTTAGGGGTGGGATTCCA 735
DB 734 SGGAGGAG-----GMAAGGGGAGGNA----- 754
QY 736 CTCGTGTGTTTCTATTTTAAAGAGACATTTTAAAGAAATGGTCAAGTTGGTGCTT 795

DB 755 -----GSLTLNASGAGGSGSSLSGKAGA-----GGAG--- 783
QY 796 CTCAGATTTCTGAGAAATTCCTTTGTATTGATATATTACATGATCATCCGACTGAGATA 855
DB 784 ---GSAGLFGSGGAGNGGYSLNGTGGD-----GTGAGGQITGLRSFGGAGG 830
QY 856 TTGTTTTACAATAG-----TTCTGTGGGCTGTGTTTTTTGTTATTAAACAAATAATTTAGA 911
DB 831 AGGASDTGAGGCGAGGKAGLYGNGDGGAG-----GDGATSGKGGAGGNAVVGNGG 883
QY 912 TGGTGAATAAAAAA 925
DB 884 NGNAGKAGGTAGA 897

RESULT 9
A70812
Hypothetical glycine-rich protein rv0833 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 07-Mar-2003
C;Accession: A70812
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70812
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-749 <COL>
A;Cross-references: GB:AL123456; NID:93261550; PIDN:CAAL17639.1; PID:92916
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv0833
C;Superfamily: elastin related uncharacterized glycine-rich protein, PE motif containi

Query Match 12.4%; Score 713; DB 2; Length 749;
Best Local Similarity 30.2%; Pred. No. 4.2e-26;
Matches 289; Conservative 21; Mismatches 361; Indels 286; Gaps 41;

QY 23 GTCCCAAGATCATGAAATCGCCAGTGGCAGCACCGCCACCGCCCGCGGCCCCAG 82
DB 3 GNGGAGSGAPG-AITGAGGPAG-LIIVGGAGGAG--GDSAVAGVIGAGGAGGALLFG 58
QY 83 CTGC-----GCCCTGAAGCCGCAAGACAGCAGCGGTGCGGGAGGTGTGCGCTGTC 138
DB 59 AGGAGAGSGSAGAGGAGGAGGAGGLFASG--GSGFGGFPASTGTGAGGTG--GAGGLP 116
QY 139 TGTCTGACAGAGCTGCGCATCTCGCGTGCCTG-----GGCGCCGGCGCGCTGCT 194
DB 117 AS--CGVGTGGGAGSG-----TGCVGTGGAGGLFASGAGGAGGSG--GTGAGG--- 165
QY 195 GCCCTGCTGGACGAGCA-----GCAGTAA--ACGTGCTCTCTACGACATG 239
DB 166 ---TGGAGLFGAGAGLGGGQNHGTGGHAGGAGSAGLLALGDDGAG---GAGAAATTG 218
QY 240 -AACGGCTGTTTACTCAAGCTCAAGAGCTGTGCCCACTCCCGCCAGAACCCAGAG-- 296
DB 219 TGGAGAGGKAGLLFGSGGAGSGGAAAGTGGTGTG-----NSGGAGGAGGKAGLL 267
QY 297 -GTGAGCAAGTGGAGATTCTCCAGCAGCTCATCGACTACATCAGGACCTTCAGTTGA 355
DB 288 FESG---GAGSGGAG-----GFANGSTGGAG-----AGGGA---GLTNGNG 306
QY 356 GCTGAACCTCGGAATCCGAAGTTGGGACCCCGGGGGCGGAGGGTGGCGGTCCG--GCTC 414
DB 307 GSGGTSVATGAGN--GGAGGAGGAGLIGNGGG---SGMGDAPGGTGVGGIGLL 360
QY 415 CGCTCAGCACCTT-----CAACGGCGAGA-----TC 440

A;Residues: 1154-1409, 'A', 1411-1420, 'K', 1422-1425, 'R', 1427-1481 <ADA3>
A;Cross-references: EMBL:X06741; NID:99355; PID:9929603
R;Yang, Y.; Adam, R.D.

Nucleic Acids Res. 22, 2102-2108, 1994

A;Title: Allele-specific expression of a variant-specific surface protein (VSP) of Giardia
A;Reference number: 548056; MUID:94301794; PMID:8029018

A;Accession: S48056

A;Molecule type: DNA

A;Residues: 1-56 <YAN>

A;Cross-references: EMBL:125059

A;Experimental source: trophozoites WBA6

A;Note: the source is designated as Giardia intestinalis

C;Comment: This translation was produced by PIR staff from information provided by the a

C;Genetics:

A;Gene: VSPA6

C;Keywords: surface antigen; tandem repeat

Query Match 12.6%; Score 725.5; DB 2; Length 1766;

Best Local Similarity 22.8%; Pred. No. 1.9e-26;

Matches 370; Conservative 23; Mismatches 461; Indels 767; Gaps 92;

QY 3 GCGCCATCTCTTT TCAGC-----CAGTCG-----CAGTAA-----33

DB 186 GTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTN 245

QY 34 -----TCAT-----GAAA-----GTCCGCACTGG-CAGC-----56

DB 246 PSDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPY 305

QY 57 -----ACGCCACCCCG-----CCGGGGGGCCCG-----CTGGCG 89

DB 306 LKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCT--KCD 363

QY 90 CTGAA-----GGC-----CGSCA-----AGACAGCGCGGTGGC--120

DB 364 ANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPC--TACAGTA 421

QY 121 -GCAGAGTGGCGCTGTC-----TGCTGA-GCAG-AGCGTG-----156

DB 422 DKC-----TKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECK 474

QY 157 CCATCTCGC-----CTGCGGGGC-----GCCGGGC-----184

DB 475 CNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDA 534

QY 185 -GCGCTCGCTGCC-----CTGCTG-----GAC--GAGCAGCAG-----215

DB 535 KECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYTTDD 594

QY 216 -----GTAACGCTGCTCT-----CTAC-----CAGATGAACGCTG 247

DB 595 SVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAG 654

QY 248 -----TTACTCAG-----CCTCAAGGAG-----C 267

DB 655 YTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDC 714

QY 268 TGGTGCC-----CACCTGCCCCAGAA-----CCGCAAG-----TGAGC 302

DB 715 QGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTC 773

QY 303 --AAGGTGAGATT-----CTCCAGCAGCTCATC-----329

DB 774 VSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTA-DKCTKCDANGAAPYLKKTNP 832

QY 330 ---GACTACATCAGG-----GACCTCAGTTGGAGCT-----GAA--361

DB 833 DPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTAD--KCTKCDANGAAPY 890

QY 362 -----CTCGAAATCCGAG-----TTGGAGCCCCGGGGCGGAGGCTGCGG-404

DB 891 LKKTNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDAN 950

RESULT 8

H70987

hypothetical glycine-rich protein Rv1759c - Mycobacterium tuberculosis (strain H37RV)

N;Alternate names: wac22 antigen homolog

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C;Accession: H70987

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, N.;

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

QY 859 ---TTTACATAGTCTGTGGGCTGTTTTTTTGTATTATAAACAATATTAGATGGT 915
Db 2244 GSDSAAAAAASAGAGSGYGGYGSDSAAAAAASAGAGSGYGGY 2303
QY 916 G-----AAAAAASAA 926
Db 2304 GYGSGYGSASAAAAAASAA 2323

RESULT 3
F70806
hypothetical glycine-rich protein Rv3508 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70806
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70806
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1901 <COL>
A: Cross-references: GB:AL022022; GB:AL123456; NID: G3261554; PIDN: CAA17745.1; PID: G29244
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: Rv3508
C: Superfamily: collagen alpha 1(IV) chain

Query Match 14.6%; Score 841; DB 2; Length 1901;
Best Local Similarity 32.3%; Pred. No. 1.1e-31;
Matches 335; Conservative 13; Mismatches 463; Indels 226; Gaps 40;

QY 1 GGGGCCCATCTCTTTCAGCCAGTCGCCAAGATCATGAAAGTCGCCAGTGGCAGCACCG 60
Db 180 GAGGWLFGVGGAGVGAG--GGTGGAGGPGGLTWGGGAGGCGGAGGTGGAGGRA--E 235
QY 61 CCACCGCCGCGCGGCCAGCTCGCGCTGAAGCGCGGCAAGACAGCG--AGCGGTGC 118
Db 236 LLFGAGGAGGAGTGG--PQATG-GTGGHGGVGGDGLWLPAGGAGGAGGCGGAGGAGS 290
QY 119 GGGC--GAGTGGT---GCCTGTC---TGCTGACAGAGCTGCCCATCTC-GGCT 168
Db 291 DGGALGGTGGTGGAGGAGGALLLGGAGGCGGAGGCGGTGGAGGAGGAGGAGGAGG 350
QY 169 GCGCG-GGCGCGCGCGCGCG--CCTGCTGCTGCTGCTGAGCAGCAGCAGTAAACCTGC 225
Db 351 GKGGGVGVAGLGGAGGAAGQLFSGAGGAAGAVGVGGTGGGAGGAGGAGGAGGAGG 409
QY 226 TGCTTACGACATG-----AACGGCTGTACTCAGCCTCAGGAGCTGTGCCACCT 280
Db 410 TGGTGFAGGAGVGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 455
QY 281 GCGCCAGAACCGAAGGTGAGCAGGTGAGATCTCCAGCAGCAGTCCATCCACTACATCAG 340
Db 456 GADNAGSAGTADG-GAGGTGAGAGGAGGAGGAGT-----GGTGGVGAAGKAGIGGTGG 508
QY 341 GGA-----CCTTCAGTTGAGCTGAAGTCCGATCCGAGTGG-----GACCCCGGG 389
Db 509 GGAGGAGSAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 563
QY 390 GSCCCAGG-----CTGCGCGTCCGGCTCCGCTCAGCAGCAGTCCAGCAGGAGATC 440
Db 564 GAGAGAGGAGVADNPTGIGTGGTGGKAGAGGAGG-----GSSGAGGTN 608
QY 441 AGCGCCCTGACCGCCGAGCG-----GCATGCGTTCTCGGAGCAG-TGCGATCTTGT 492
Db 609 GSGGAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 663
QY 493 -CTGCTGAAGCGCTCCCGCAGGAGCGCGGGA-CCCCAGCCATCCAGGGGCGCAAGAG 550

Db 664 GAGSVGNAGIG-----GTGGTGGVGGAGGAGAAAAAGSSAT-----CGAGFAGGAGG 712
QY 551 AATACGTGCTCTGTGGTCTCCCCCAACGCGCTCCCGGATCTGAGGGAGAACAGAC 610
Db 713 EGGAGNSGVGTNGSG-----AGGAGGKGTGGAGGAGADNPTGAG 755
QY 611 CGATCGCGCGCCACTGCGCCCTTAACATGC--ATCCAGC-----CTGGGGCT----- 654
Db 756 FAGGAGGTGGAGAGGAGGAGTCTGGTGGVGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 815
QY 655 ----GAGGCTGAGGACTGGC-GAGGAGAGGCGCTCTCTCTGCACACCTACTAGTAC 709
Db 816 GFDGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 864
QY 710 CAGAGACTTTAGGGGCTGGGATTCCTCCTGCTGTGTTCTATTATTTTGAAGAGCAG----- 764
Db 865 VGDDGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 915
QY 765 -----ACATTTAAAAATGTCAGCTTTGGTGTCTTCTCAATTT----- 804
Db 916 GCGGAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 975
QY 805 -----CTGAGGAAAT-----GCTTTGTATTGTATATTACATGATCACC 844
Db 976 DAGSGGGGFGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1034
QY 845 GACTGAGATATTTTACATAGTCTG-----TGGGGCTGTTTTTTTGTATTAAAC 899
Db 1035 GGINGAG 1094
QY 900 AATATATTAGATGGT 916
Db 1095 AAGGNGGAGGAGGAG 1111

RESULT 4
B70807
hypothetical glycine-rich protein Rv3512 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70807
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1079 <COL>
A: Cross-references: GB:AL022022; GB:AL123456; NID: G3261554; PIDN: CAA17749.1; PID: G29244
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: Rv3512
C: Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 14.2%; Score 819; DB 2; Length 1079;
Best Local Similarity 31.9%; Pred. No. 8.1e-31;
Matches 340; Conservative 16; Mismatches 462; Indels 249; Gaps 38;

QY 2 GGGCCCATCTCTTTTTCAGCCAGTCCGCAAG-AATCATGAAAGT--CGCCAGTGGCAGCAC 58
Db 106 GTGSAFPTAGTGGD---GGKGGNGGTGAAGTTGPVGTGASGGTGGGAGGAGGAGGAGGAGGAG 162
QY 59 GCGCCAGCGCGCGCGGG-----CCCCAGCTCGCGCTGAAGCGCGGCAAGACAGC----- 109
Db 163 GGTAGAG 221
QY 110 -GAGCGCGTGG-----GCGAGTGTGCTGTCTGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 160

Qy 913 G-----GTGAAAAAAA 925
Db 1519 GCAGGCCATGTGCACGAGACA 1539
Search completed: May 7, 2004, 15:25:49
Job time : 73 secs

Qy 219 AACGTGCTGCTACGACATGAACGGCTGTTACT-----CAQGCCTCAAGG-- 264
Db 462 CGCGCCCTGGCT-----CATGACGGGGCTCCCGCTGGCGGGGGCGCCCGGGCT 517
Qy 265 -----AGCTGGTGC-----CACCTGGCC-CAGAACCGCAAGGTGACGAGG 306
Db 518 GTGAATCGGACTCGCCCTCGCGGGCTCCCGCCCGCCCGCCCGCGG---ACGTGG 574
Qy 307 TGGAGATTCACGACGCTCATCGA-----CTACA-----TCAGGG-- 342
Db 575 TAGGGGATGCCAGCTCCACTGCGATGGAGTTGGCGGCTCTCCAGTTCCCTCTGCTC 634
Qy 343 ACCTTCAGT-----TGGAGCTGAAT---CG-GAATCC-----GAAGTTGGACCC 384
Db 635 ACCTGCTGCTGATGGTGTCTGTGCACTCCGAGCATCCGCTGGAGAAGCTGG---C 690
Qy 385 CCGGGGCGGAGGCTCGCGGTCCGG-----CTCCGCTCAGCA---CCCTC 428
Db 691 CCAGCCACC-AGAGCAGCGGCGCAGGAGAGCGTGAGCACGCACTCGGGACGGCCCG 749
Qy 429 AACGG-----CGAGATC-AGCGGCC--TGACGGCCGA-GCGCGCATGCGTTC---CT-- 473
Db 750 GCGGGGTGAACGAGCTCGGGCGCGCGGAGGACGAGGGCGGCA-GCGGCCGGGACTGG 808
Qy 474 -----GCGGAC-----GATCGC--ATCTTGTGTC-----GCTGAGC--GCC 506
Db 809 AAGAGCAAGAGCGCCCTGGGCTCGCGCGCTGAGCCGTGGAGCAAGCTGAAGCAGGCC 868
Qy 507 ----TCCCCCAGGG-----ACC--GGC--GGACC-----CCAGCCATCCAGGGG-- 542
Db 869 TGGGTCTCCAGGGCGGGGGCGCAAGCGGGGATCTGCAAGTCCGGCCCGCGGGAC 928
Qy 543 -----GCAAGAGGAATTACG-----TGCTGTG-----TGGGTCTC-CC 574
Db 929 ACCCGCAGCGGGAAGCCCTGGCGCGAGCCCGCCAGGACGCGATTTGCCCCGGAACTCGG 988
Qy 575 CCAACGC-----GCCTC-----GCCGA--TCTGAGG-----GAG 602
Db 989 CCCAGCCCGAGCCACCGAGGAGTACGTGTACCCGGACTACGCTGSCAAGGGCTGGTG 1048
Qy 603 AACAG-----AC-CGATCG-----CG-----CCACTGC--GC 629
Db 1049 GACGAGCGGCTTGTGTACGCGATCGGGAGAGTTCCGGCGCGGCCCTCGGCTGC 1108
Qy 630 CCTTAAC--TGCATC-----CAGCCTGGGG--CTGAGGCTGAG-- 663
Db 1109 CCGTCCCTGTGACCGAGAGGGGCGCTGTGCGCGCAGCCCGAGTCCCGAGGCTGCAC 1168
Qy 664 --GCACT-----GGCGAGGAG--GGCGCT--CCTCTCTCC-----AC 696
Db 1169 CCGCGCTGCATCCAGTCCGACAGCAGCCAGTCTGCCCGCAGTCAAGAGAGAGAGAAC 1228
Qy 697 ACCTACTAGTCAACG-----AGACTTTAGGG-----GTGG-GATTCCACT 737
Db 1229 TACTCGAGT--TCCGGGCAAGACCTATCAGACTTTGGAGGAGTTCGTGTCTCCATG 1287
Qy 738 CGT---GTGTTTCTATT---TTTGAAG---CAGACATTTTAA 774
Db 1288 CGAGAGGTGTCGTGTGAAGCCACCGTGAGGTGTATGACAGTGTGAG-CGTGTCCC 1346
Qy 775 AAA-----TGTCTAC-----GTTTGTGCT-TCTCAGATTCTGAGGA 811
Db 1347 AGACGAGTGTGTGACCCCTGTGTACGAGCCGTGATCAGTGTCTCCCA----TCTGCAA 1402
Qy 812 AAT-----TGCTTTGTATTGTATTATTAATGATC---ACCGACTGAGAA-TATTG 858
Db 1403 AATGTCCAAACCTGCTTTGCA---GAAACCGCGGTGATCCCTGTGGCAGAGAGTGAAG 1459
Qy 859 TTTTACAATAGTTC--TGTGGGCTGTTTTTTTGTATTATTAACAAT---AATTAGAT 912
Db 1460 ACTGACGAGTGACCATATGCACTGTATG-AGGAGGACATGGAGAAATCGAGCG 1518

Db 402 AACCCGGGCGCGCGCGCTGCTACCCCTGGCGCCCTGGAGCCGCGCTCGGCC 461
Qy 219 AACGTGCTGTACACATGAACCGCTGTACT-----CAGCCTCAAGG-- 264
Db 462 CGCGCCCTGGCT-----CATGGACGCGCTCCCGCTGGCGCGCGCGCGCT 517
Qy 265 -----AGCTGGTGC-----CACCTGCC-CAGAACCGCAAGGTGAGCAAG 306
Db 518 GTGAATGCGACTCGCCCTCGCGCGCTCCCGCGCGCGCGCGCGCGCGG--ACGTGG 574
Qy 307 TGGAGATTCTCCAGCAGCTCATCA-----CTACA-----TCAGGG-- 342
Db 575 TAGGGATGCCAGCTCCACTGCGATGGCAGTTGGCGGCTCTCCAGTTCCCTCTGGTTC 634
Qy 343 ACCTTCAGT-----TGGAGCTGAAT--CG-GAATCC-----GAAGTTGGACCC 384
Db 635 ACCTGCTGCTGATGTGGTCTGTGATCCGATCCGATCCGATGGAGAGCTGG--C 690
Qy 385 CCGGGGCGCGAGGGCTGCGGCTCGGG-----CTCGGTCAGCA--CCCTC 428
Db 691 CCAGGCACC-AGAGCAGCGCGCGCAGAGAGCGTGAGCAGCCTCGGGAGCGCGCG 749
Qy 429 AACGG-----CGGATC-ACGGCC--TGACGGCGA-GGCGCATGCGTTC--CT-- 473
Db 750 GCGGGGTGAACAGCTCGGGCGCGCGCGAGAGCGAGGCGGCA-GCGCGCGGACTGG 808
Qy 474 -----CGCGAC-----GATGCG--ATCTTGTGTC-----GCTGAAGC--GCC 506
Db 809 AAGAGCAAGAGCGCGCTGGCTCGCGCGCTGAGCGGTGGAGCAAGCTGAAGCAGGCG 868
Qy 507 -----TCCCCCAGG-----ACC--CGC--GGACC-----CCAGCATCCAGGG- 542
Db 869 TGGGTCTCCAGGCGCGGGCGCGCAAGGCGGGGATCTGAGTTCGGGCGCGCGGGAC 928
Qy 543 -----GCAAGAGGAAATACG-----TGCTCTG-----TGGGTCTC-CC 574
Db 929 ACCCGGAGGCGGAGCCTGGCGCGAGCGCGCGCGAGCGATGGCGCGGAATCGCG 988
Qy 575 CCAAGC-----GCCTC-----GCCGA--TCTGAG-----GAG 602
Db 989 CCCACGCGCGACCGAGGAGTACGTGTACCGGACTACCGTGCCAAAGGCTCGGT 1048
Qy 603 AACAG-----AC-CCATCG-----CGG--CCACTGC--GC 629
Db 1049 GACGAGCGGCTTGTGTACCGATCGGGAGAGTTGCGCGCGCGCGCTCGGCTGC 1108
Qy 630 CTTAAC--TGCATC-----CAGCTGGG--CTGAGGCTGAG-- 663
Db 1109 CCGTGCCTGTGCACCGAGGAGGCGCGCTGTGCGCGCAGCGCGAGTCCCGAGGCTGCAC 1168
Qy 664 --GCACT-----GGCGAGGAG--GGCGCT--CCTCTGTC-----AC 696
Db 1169 CCGGCTGCATCCAGCTGCACACGAGCAGTGTGCGCGCGAGTCCAGGAGAGAGAAC 1228
Qy 697 ACCTACTAGTCACCA-----AGACTTTAGGG-----GTGG-GAATCACT 737
Db 1229 TACTGCGAGT--TCCGGGCAAGACCTATCAGACTTTGGAGAGTTCTGTGTCTCCATG 1287
Qy 738 CGT--GTGTTCTATTT-----TTTGAAG--CAGACATTTTAA 774
Db 1288 CGAGAGGTGTCTGTGAAGCAACAGGTGAGGTGCTATGCACTGTCTGCTCC 1346
Qy 775 AAA-----TGCTCAC-----GTTTGTGCT-TCTCAGATTTCTGAG 811
Db 1347 AGACGGAGTGTGTGACCCCTGTGTACGAGCCTGATCAGTGTCTGCCA-----TCTGCAA 1402
Qy 812 AAT-----TGTTTGTATGTATATTACATGATC-----ACGACTGAGAT-TATTG 858
Db 1403 AATGTTCCAACTGCTTTGCA--GAAACCGCGGTGATCCCTGTGCGAGAGAGTGAAG 1459
Qy 859 TTTTACATAGTTC--TGTGGGCTGTGTTTTTTTGTATTAACAAAT-----AATTTAGAT 912

Db 1460 ACTGACGAGTGCACCATATGCCACTGTACTTATG-AGGAAGGCACATGGAGATCGAGCG 1518
Qy 913 G-----GTGAAAAAAA 925
Db 1519 GCAGGCCATGTGCACGAGACA 1539
RESULT 15
US-10-123-155-515
; Sequence 515, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 515
; LENGTH: 1942
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-515
Query Match 47.6%; Score 2744.5; DB 14; Length 1942;
Best Local Similarity 41.1%; Pred. No. 1.5e-129;
Matches 625; Conservative 0; Mismatches 275; Indels 621; Gaps 96;
Qy 3 GSCC-----CAT-TCTGTTTCAGCC-AGTCCCAAGAAT--CATGAAGTTC--- 44
Db 42 GGCCTGGCTCCGGGCTTCTTTTCCCTCCGACGCGCCACGCTGCCACGACATTCCGCGC 101
Qy 45 -GCCAG--TGGC-AGCACCGCCACCGCGCGCGG-- 74
Db 102 TGCCGGGTCTGGAGAGCTCCCGAACCCCTCCGCGAGAGAGCGAGCGCGCAGGCT 161
Qy 75 GGCCTCCGCTGGCG-- 92
Db 162 GGCCTCCGCGCGCGCTTGGTCTCGAGAGCGGGGAGCGAGCGGAGTACGAGCTG 221
Qy 93 AAGG-----CCGCA--AGACAGC----- 109
Db 222 AGGCGCAGCGCGGCACTGACGCGAGTTGGGCGCGGCTACCGGAGCTGACAGCGCAT 281
Qy 110 GAGCG-----GTGCGGCGAGGTGGTGGCG--TGCTGTG 140
Db 282 GAGCGACTCCCGAGAGCGCCCTAGCCGTTGTGCGCGAGCGGAGCGCGCAGTGGG 341
Qy 141 TCTGAGCAGAGCGTGG-CCATCTCGCG-----TGCGG-----GGCG-- 177
Db 342 GCTGGGCTTGTAGTGTCCGCCACCGGGGTGCGCGCGCGCGCGCGCGCGTGGC 401
Qy 178 ----CGGG--GCGGCGCTG-----CCTGC--CTTGTG--GACGAGCAGAGTGA 218
Db 402 AACCCGGGCGCGCGCGCGCTGTACTACCCCTTGGCGCGCGCTCGGAGCGCGCGCTCGGCC 461

Db 2020 GCGGTTGGCGGTGGCGCCGGGTTCCCGGAGGCGCTGCTGGCGCCCTGATGAGAAC 2079
QY 211 AGCAGATAAA---CGTGTGCT---CTACGACATGAAC---GGCTGTGA 250
Db 2080 GGTGGGCCAAGCCACGCTGCTGACGGGGGCCCCACGACTGGACTCGGGGTGCTG 2139
QY 251 CTCAGCCT---CA---AGGAGCTGGTCCCAACCTGCCCCAGAACGCG 293
Db 2140 CCCACCCCGAGCAGACGCGCTGCCGCGAAGCGCTGCCCACTCCGACCCGACGCC 2199
QY 294 AAGG---TGAGCAAGGTG---GAGATTCTCCAG---CACGTTCATCGACTACA 336
Db 2200 CAGCCCTGGGCCCCCGCGCTGGGACACAGGCCACCCCTGCTCCGCGCTCCGCTTCA 2259
QY 337 TCAG-GGACCTTCAGTTGAGCTGAACCTCGGAATCCGAAG---TTGG--- 379
Db 2260 TCCTCCCTCTGCTGCTGGCGC-CGCCCGGGCCCCCGAGCAGCCCCCGCGCTGGGA 2318
QY 380 ---GACCCCGGGGCG---AGGCTG-CGGTCCGGGC----- 412
Db 2319 GCGGACCCCGAGCGCGCGCTCTATGCTCCCGCGCCCGCGCGCTCCACGGGACTT 2378
QY 413 TCGCTCAGCACCT---CAAGCGGAG---ATCAGCGCCCTGACGG-- 453
Db 2379 CCGCTCACCCCCACGCGCCGACCGCGCGGGTGGTTCGGCGCC-CACGGC 2436
QY 454 -CGAGS---CGGATGCGTCTCTGCGGAGATCGCATCTTGTGCTGAAGCGCTCC 509
Db 2437 CCCTTGACCCAGCCT---CAGCGCGGATGGCTCCCGCGCCCTGGAGC-CGCGC 2489
QY 510 CCC--AGGG--ACC---GGC---GGACCCCA---GCCATCCAG-G 540
Db 2490 CCCGAGCGGAGCTGAGGAGCCACTGGGCCCCCAGCGCCCTCGGCGCCACCTGG 2549
QY 541 GGGCA-----AGAGGAATTAGTCTCTGTGGTCTCCCCCAAGC-GCCT 585
Db 2550 CCGCACCCACACGTTCAACAGCGGAGCGCGCCCTGGGACCGCACCGGGGTGCA 2609
QY 586 CGCC---GGATCTGA-----GGGAGAACAGCCGATCGG 617
Db 2610 CGCCCGCGGCGACAGACTTGGCCACCTCTCCCTATGGGGGC-GGACAGACTG 2668
QY 618 CGGCCACTCGCCCTTAAC-----TGCA-----CCAGCTGGGG--- 652
Db 2669 CGCCCGCGTGCCTTAGCGCGGGGCCCGCGATGCTTGGCAGTCCACGACCGGAA 2728
QY 653 -----CTGAG-GCTGAGGCACTGG----- 670
Db 2729 CAGGAGCAGAGCGGTGCCAGAACCGCGGGGCCCGGGCACTCCGAGTGGTGTCTAA 2788
QY 671 -----CGAGGAGGGCGCTCCTCTC-----TGCA-ACC 699
Db 2789 GTCCCCCGGACCCACCGCGAGTGGGGGCCCGCTCCGCGCACAGGAGCAACACC 2848
QY 700 TACTAGTC-----ACGAGACT-----TTAGGGGTG----- 727
Db 2849 AGCTCGCCCTCCCTTACCCGGGCGCGAGACGCTGAGACGGTTTGGGGGTGGGTCG 2908
QY 728 -----GGATT-----CC---ACTCGTGTGTT---CTATTTT 755
Db 2909 GGGAGGACTTGCTATGATTTGAGTGTGACCTTATGCGGTAGGTTTGGTTTTTTT 2968
QY 756 GAA-----AAGCAGACATTTTAAATGTTCA-----CGTT-----TGCT 791
Db 2969 GCAGTTTGGTTTCTTTTGGGGTTTCTAACCAATTGCACACTCCGTTCTCGGGGTGGC 3028
QY 792 -----GCTT---CTCAGAT-----TTCTGAGGAATTG- 816
Db 3029 GGCAGGCGGGAGGCTTGGACCGCGGTGGGATGGGGGCCACAGCTGCAGACCTAG 3088
QY 817 -----CTTGTATGTATAT-----TACAATGATCACGAC-----TGAGATAT- 856

Db 3089 CCCTCCCCCACCCTGGAAAGGTCCCTCCCAACCCAGGCCCTGGCGTGTGGGTGTG 3148
QY 857 -----TGTTT---TACAAAGTTCGTG---GGCTGTGTTTTTTT 892
Db 3149 CGTGGTGTGCGTCCGCTGTTGCTGTAAGGGCGCGGAGGTGGGC-GTGTGTGTGCG 3207
QY 893 ATTAAACAATAATTATAGATGGTGAATAAAAAA 926
Db 3208 TGCCAGCAAGGCTGCTGTGGGCGTGTGTGTCAA 3241

RESULT 13

US-10-140-864-543
; Sequence 543, Application US/10140864
; Publication No. US20030207419A1
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C184
; CURRENT APPLICATION NUMBER: US/10/140,864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-864-543

Query Match 48.4%; Score 2790.5; DB 15; Length 3721;

Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;

QY 1 GGGGCC-----CATTC-----GTTTC-----AGCAG--TCGCCAAG 31
Db 1723 GGGGCCAGCACCTCAGGCTTAGGGGACTGCACAGGACTCCTGCGGCGCAGCTCTCCGAG 1782
QY 32 AATCATG-----AAAGTCGCCAGTGGCAGC----- 56
Db 1783 GACCGCGGGGCTGGTGTGCGTGAACCTGCTGTACGTCTGCGTGGCGCCTTCGTG 1842
QY 57 -----ACCG-----CC-ACCGCGC-----CGCGGCC-CC---AGCTGC-- 86
Db 1843 GTGGAGCGGTGTGCTCGGCTTCAGCGTGGCTGTTCTGTGGCTCCCTGAGCGCGG 1902
QY 87 CGCTGAGGCGCGC---AAGACA-----GGAGCGGTGCGGCGAGTGG 129
Db 1903 GAGCTG--GCCCGGCGAGGAGCAAGAGGCCATCTTGGCGCAGCGGCGCGGAGCGG 1960
QY 130 TGC---GCTGTCT---GTCTGAGC-AGAGCGTGGC-CA---TCTC----- 163
Db 1961 TGCTGAGC-GTCAGCGCTGGCGAGCGCAGGCGCAGAGGTCCCGGGGCGCGGCGGA 2019
QY 164 -CGCTGCGCG- GCGCGCGGGCGCGCTGCTGCTGCTGCTG-----GACGAGC 210
Db 2020 GCGGTGGCGTGGCGCGGGGTTCCCGGAGGCCCTGCTGGCGCCCTGATCGAGAAC 2079

QY	817	-----	CTTTGATTGTATAT	-----	TACAATGATCACCAGC	-----	TGAGAAATAT	856
DB	3089	CCCTCCCCAC	CCCTGGAAAGTCCCTCCCA	CCCAAGCCCTCGCGTGTGGTGTG	3148			
QY	857	-----	TGTTTT	-----	TACAATAGTTCTGTG	-----	GGCGTGTGTTTTTCTT	892
DB	3149	CGTGC	GTGCGTGC	CGCGTTCGTGTGCAAGGGCCGGGAGTGCGC	GTGTGTGTCG	3207		
QY	893	ATTAAACAAATAAT	TTAGATG	TGTGAAAAA	AAAAA	926		
DB	3208	TCCCAGCAAGGCT	GCTGCTGGCGCTGTGTGCA	3241				

130	QY	TGC---GCTGCTCT---GCTGTAGC-AGAGCGTGGC-CA---TCTC-----	163
131			
132			
133			
134			
135			
136			
137			
138			
139			
140			
141			
142			
143			
144			
145			
146			
147			
148			
149			
150			
151			
152			
153			
154			
155			
156			
157			
158			
159			
160			
161	DB	TGCTGAGC-GTCAGCGCGCTGTGGCGAGCGCAGCGCGCAGGGTCCCGGGGGCCGGGGCGGA	2019
162			
163			
164	QY	-CGGCTTGGCGG-GGCGCGGGGCGCGCTGTGCTGCCCTGCTG------GACGAGC	210
165			
166			
167			
168			
169			
170			
171			
172			
173			
174			
175			
176			
177			
178			
179			
180			
181			
182			
183			
184			
185			
186			
187			
188			
189			
190			
191			
192			
193			
194			
195			
196			
197			
198			
199			
200	DB	GCTGGGCCAAGGCCACGCTGTGTGAGGGCGGGGCTCCACGACCTGGATCTCGGGGCTGTG	2139
201			
202			
203			
204			
205			
206			
207			
208			
209			
210			
211	QY	CTCAGCCCT---CA-----AGGAGCTGGTGCCACACCTGCCCCAGAACCGC	293
212			
213			
214			
215			
216			
217			
218			
219			
220			
221			
222			
223			
224			
225			
226			
227			
228			
229			
230			
231			
232			
233			
234			
235			
236			
237			
238			
239			
240			
241			
242			
243			
244			
245			
246			
247			
248			
249			
250			
251			
252			
253			
254			
255			
256			
257			
258			
259			
260			
261			
262			
263			
264			

QY	792	-----GCCTT---CTCAGAT-----TTCTGAGGAATTG- 816
Db	3029	GGCAGGACGAGGAGGCTTTGGAGCGCGTGGGGAATGGGGGGCCACAGCTGCAGACCTAAG 3089
QY	817	-----CITTGPTATTGTATAT-----TACAATGATCACCAGC---TGAGAATAT- 856
Db	3089	CCCTCCCCACCCCTCTGGAAGGTCCCTCCCAACCCAGGCCCCCTGGCGTGTGTGGGTGTG 3148
QY	857	-----TGTTT---TACAATAGTCTCTGTG---GGGCTGTTTTTTTGT 892
Db	3149	CGTGGGTGTGCGTGCCCGTGTTCGTGTGCAAGGGCCGGGAGGTGGGC-GTGTGTGTGG 3207
QY	893	ATTAAACAAATAATTAGATGTTGAAAAA 926
Db	3208	TGCCAGCAAGGCTGCTGTGGGCGTGTGTCAA 3241
RESULT 10		
US-10-141-756-543		
; Sequence 543, Application US/10141756		
; Publication NO. US20030207359A1		
; GENERAL INFORMATION:		
; APPLICANT: Baker, Kevin P.		
; APPLICANT: Beresini, Maureen		
; APPLICANT: DeForge, Laura		
; APPLICANT: Desnoyers, Luc		
; APPLICANT: Filvaroff, Ellen		
; APPLICANT: Gao, Wei-Qiang		
; APPLICANT: Gerritsen, Mary E.		
; APPLICANT: Goddard, Audrey		
; APPLICANT: Godowski, Paul J.		
; APPLICANT: Gurney, Austin L.		
; APPLICANT: Sherwood, Steven		
; APPLICANT: Smith, Victoria		
; APPLICANT: Stewart, Timothy A.		
; APPLICANT: Tumas, Daniel		
; APPLICANT: Watanabe, Colin K		
; APPLICANT: Wood, William		
; APPLICANT: Zhang, Zemin		
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC		
; FILE REFERENCE: P3330R1C200		
; CURRENT APPLICATION NUMBER: US/10/141,756		
; CURRENT FILING DATE: 2002-05-08		
; Prior Application removed - See File Wrapper or Palm		
; NUMBER OF SEQ ID NOS: 550		
; SEQ ID NO 543		
; LENGTH: 3721		
; TYPE: DNA		
; ORGANISM: Homo Sapien		
US-10-141-756-543		
Query Match 48.4%; Score 2790.5; DB 15; Length 3721;		
Best Local Similarity 39.7%; Pred. No. 1.3e-131;		
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;		
QY	1	GGGGGCC-----CAITCT-----GTTTC-----AGCCAG--TCGCCAAG 31
Db	1723	GGGGCCAGCACCTCAGGCTTAGGGGACTGCACAGGACTCTGCGGGCCAGCCTCTCCGAG 1782
QY	32	AATCATG-----AAAGTCCCACTGGCAGC----- 56
Db	1783	GACCGGCGGGGCTGTGTGCGTGGAACCTGTGTAACGTGTCGTCGGTGGCGGCTTCGTG 1842
QY	57	-----ACCG-----CC-ACCGCCG-----CCCGGGCC-CC---AGCTGC-- 86
Db	1843	GTGGGAGCCGTGTGTGTCGCGCTTCAGCGTGGGCTGTTCTGTGGCCCTCCGTGAGCGCGG 1902
QY	87	GCCTGAAGCCGCG---AAGACA-----CCGAGCGGTGCGGCGAGGTGG 129
Db	1903	GAGCTG--GCCCGGCGCAGGACAGAGGACCATCTCGCGCACGCGGGCGGCGAGCGG 1960
QY	130	TGC---GCTGTCT---GTCTGAGC-AGACGCTGGC-CA---TCTC----- 163

```

Db 1903 GAGCTG--GCCCGGCGAAGGACAGAGAGCCATCTCTGCGCAGCGGGCGGCGAGGCGG 1960
QY 130 TGC---GCTGTCT---GTCTGAGC-AGAGCTGTC-CA---TCTC-----163
Db 1961 TGCTGAGC-GTCAGCGCTCTGGCGAGCGCAGGCGCAGAGGTCCCGGGGCGCGGCGGA 2019
QY 164 -GGCTGCGCG-GGCGCGGGCGCGCTGCGCTGCGCTGCTG-----GACGAGC 210
Db 2020 GCGGTGGCGGTGGCGCGGGTCTCCCGCGAGGCGCTCTGGCGCGCTGATGAGAAC 2079
QY 211 AGCAGGTAAA---CGTGTCTCT-----CTAGACATGAAC---GGCTGTGA 250
Db 2080 GGTGGGCGCAAGSCCAGCTGCTGACAGGCGGCGCCACGACCTGACTCGGGGTGCTG 2139
QY 251 CTCACGCT---CA-----AGGAGCTGGTCCCGCAGCTGCGCGCGCGCAGAACGCG 293
Db 2140 CCCACGCGCGAGCAGCGCGCTGCGCGAAGAGCGCTGCCACCTCCGACCCGACCC 2199
QY 294 AAGG-----TGAGCAAGGTG---GAGATTCTCCAG--CACGTCTCATGACTACA 336
Db 2200 CAGCGCTGGCGCCCGCGCTGGGACACGCGCCACCTCTGCTCCCGGCTCCGCTTCA 2259
QY 337 TCAG-GGACCTTCAGTTGAGCTGAATCGGAATCGAAG-----TTGG---379
Db 2260 TCCTCCCTCTGCTGCTGGCGCC-CGCGCGGCGCGCGCGAGCAGCGCGCGCTGGGA 2318
QY 380 ---GACCCCGCGGCGCG---AGGCTG-CGCGTCCGCGC-----412
Db 2319 GCGACCGCGCGCGCGCTCTATGCTCCCGCGCGCGCGCGCTCCACGCGGACTT 2378
QY 413 TCGCTGACGACCT-----CAAGCGGAG---ATCAGCGCGCTGACG---453
Db 2379 CCGGCTCACCGCGCGCGCGCGCGCGCGCGCGGTGGTCTCGCGCGC--ACGGCG 2436
QY 454 -CGAGG---CGCATGCGTCTCGCGGAGTCGATCTGTGTCGTCGAAGCGCTCC 509
Db 2437 CCCTTGACCGACT-----CAGCGCGATGCGCTCCGCGCGCTGAGC--CGGC 2489
QY 510 CCC--AGGG--ACC-----GGC-----GGACCGCA-----GCCATCAG-G 540
Db 2490 CCGGCGGCGAGCTGAGAGGCGCACTGGGCGCGCGCGCGCGCTCCCGCGCGCACCTGG 2549
QY 541 GGGCA-----AGAGGATAGTCTGTCGCGGAGTCGATCTGTGTCGTCGAAGCGCTCC 585
Db 2550 CCGGACCGCACGCTTCAACAGCGGAGGCGCGCGCTGGGACCGCGCGCTGCGCA 2609
QY 586 CGCC-----GGATCTGA-----GGGAGAACAGACCGGATCGG 617
Db 2610 CGCGCGCGCGCACAGACTTGGCCACCTCTCTCCCTATGCGGCGGC--GGACGAGCTG 2668
QY 618 CGGCACTGCGCGCTTAAC-----TGCA-----CGAGCTGGGG---652
Db 2669 CGCGCGCGCTGCTAGCGCGGCGCGCGCGCGCGCTGGCAGTCCGACCGACCGGGAAC 2728
QY 653 -----CTGAG-GCTGAGGCACTGG-----700
Db 2729 CAGGAGCGAGAGCGGTGCCAGAACCGCGGCGCGCGCGCGCGCGCGCTGCTCA 2788
QY 671 -----CGAGAGAGGCGCTCTCTC-----TGCAC-ACC 699
Db 2789 GTCCCCCGCGACCCACCGCGAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2848
QY 700 TACTAGTC-----ACGAGAGCT-----TTAGGGGGTG-----727
Db 2849 AGCTCGCTCTCCCGCTACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2908
QY 728 -----GGATT-----CC-----ACTGCTGTGTT---CTATTTTTT 755
Db 2909 GGGAGGACTTGTCTATGATTTGAGTTGACCTTATCGCGTAGGTTTGGTTTTTTTTT 2968
QY 756 GAA-----AAGCAGACATTTAAAGATGTCA---CGTT-----TGTT 791

```

```

Db 2969 GCAGTTTTTGGTTTCTTTTGGGTTTTCTAACCAATTGACAACTCCGTTCTCGGGGTGGC 3028
QY 792 -----GCTT---CTCAGAT-----TTCTGAGGAATTG- 816
Db 3029 GGCAGGCGAGGGAGCTTGACCGCGGTGGGGAATGGGGGCGCACAGCTGCAGACCTAAG 3088
QY 817 -----CTTTCTATTGTATAT-----TACAATGATCACCAGAC---TGAGAAATAT- 856
Db 3089 CCCTCCCCCAGCTGGAAGGTCCTCCCAACCCAGGCGCTTGGCGGTGTGGGTGTG 3148
QY 857 -----TGTTT---TACAATGATCTCTGTG---GGCTGTTTTTTTTT 892
Db 3149 CGTGGGTGTCGTCGCTGTTCTGTCGAAAGGCGCGGAGGTGGGC-CTGTGTGTGCG 3207
QY 893 ATTAAACAAATAATTTAGATGGTCAAAAAA 926
Db 3208 TGCCAGCGAAGGCTGCTGTGGCGGTGTGTCTCA 3241

RESULT 9
US-10-140-923-543
; Sequence 543, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-923-543

Query Match 48.4%; Score 2790.5; DB 15; Length 3721;
Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;

QY 1 GGGGCC-----CATTC-----GTTTC-----AGCAG--TCGCCAAG 31
Db 1723 GGGGCCAGCACTCAGGCTTAGGGAGTGCACAGACTCTCGCGGCGCAGCTCTCCGAG 1782
QY 32 AATCATG-----AAATGCCAGTGGCAGC-----56
Db 1783 GACCGCGGGGCTGGTGTGCTGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1842
QY 57 -----ACCG-----CC-ACCGCGC-----CGCGGCGC--CC---ACGTCG-- 86
Db 1843 GTGGAGCCGTGTGTGCTCGGCTTCAGCGTGGGCTGGTGTGTTGGGCTCCCGTACGCGCGG 1902
QY 87 GCGCTGAAGGCGCGC---AAGACA-----GCCAGCGGTGCGGCGGAGGTGG 129
Db 1903 GAGCTG--GCCCGCGCAAGCAAGGAGGAGGCACTCTGCGCGCACGGGCGGCGGAGGCGG 1960

```

```

QY 87 GCGCTGAAGCGCGCG---AAGACA-----GCGAGCGGTGCGGGCGAGGTGG 129
Db 1903 GAGCTG--GCCCGCGCAAGGACAGAGAGCCATCTGCGCGCACCGGGCGGCGAGGCGG 1960
QY 130 TGC---GCTGTCT---GTCTGAGC-AGAGCGTGGC-CA---TCTC-----163
Db 1961 TGTGAGC-GTCAGCGCTTGGCGAGCGAGGCGCAGGCTCCCGGGGCGCGGGCGGA 2019
QY 164 -GCGCTGCCGG-GGCGCGGGGGCGCGCTGCTGCTGCTGCTG-----GACGAGC 210
Db 2020 GCGGCTGGCGGTGGCGCGGGGTTCCCGGAGGCGCTGCTGCGCGCTGATGAGAAC 2079
QY 211 AGCAGGTAA---CGTGTGCT-----CTAGCATGAAC-----GGCTGTGA 250
Db 2080 GGTGGGCGAGCGCAGCTGCTGAGGGCGGCGCCACGACCTGACTCGGGCTGCTG 2139
QY 251 CTCACGCT---CA-----AGGAGCTGGTGGCCACCTGCGCCAGAACCGC 293
Db 2140 CCCAGCGCGAGCAGCGCGCTGCGCGAGAGAGCGCTGCCACTCGCAGCCGACGCC 2199
QY 294 AAGG-----TGAGCAAGTG---GAGATTCTCAG--CACGTGATCGACTACA 336
Db 2200 CAGCCCTGGGCGCGCGCTGGAGCAGCGCCACCGCTGCTGCTGCGGCTCCGCTTCA 2259
QY 337 TCAG-GGACCTTCAGTTGAGCTGAATCGGAATCCGAAG-----TTGG---379
Db 2260 TCCTCCCTCTGCTGCTGGCGCC-CGCGCGGCGCGCGAGCAGCCCGCGCTGGGA 2318
QY 380 ---GACCGCGGGGCGG---AGGCTG-CGCGTGGCGG-----412
Db 2319 GCCAGCCCGGCGCGCTTATGCTGCGCGCGCGCGCGCGCTCCGACGGGACTT 2378
QY 413 TCCGCTCAGCACCTT-----CAACGCGGAG-----ATCAGCGCCCTGACGG--453
Db 2379 CCCGCTACCCCCACGCCAGCCGCGCGCGCGGCTGGTGTGCGCGGCC--ACGGG 2436
QY 454 -CGGAGG---CGGATGCTTCCGCGAGATCGCATCTGTGCTGCTGAAGCGCTCC 509
Db 2437 CCCTTGACCGAGCT-----CAGCGCGGATGCGCTCCCGCGCGCTGGAGC-CGCG 2489
QY 510 CCC--AGG--ACC-----GGC-----GACCCCA-----GCCATCCAG-G 540
Db 2490 CCCAGCGGAGCTGAGGAGCCACTGCGGCGCGCGCGCGCGCGCGCGCGCGCG 2549
QY 541 GGGCA-----AGAGAAATACGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 585
Db 2550 CCGCACCACACAGTTCAACAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2609
QY 586 CGCC-----GATCTGA-----GGGAGAAAGACCGGATCGG 617
Db 2610 GCGCGCGCGGACAGACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2668
QY 618 CGGCACTGCGCCCTTAA-----TGCA-----CCAGCTGGGG---552
Db 2669 CGCGCGCGCGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2728
QY 653 -----CTGAG-GCTGAGGCACTG-----670
Db 2729 CAGGAGCGAGACGCTGCCAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2788
QY 671 -----CGAGGAGAGGCGCTCTCTC-----TGAC-ACC 699
Db 2789 GTCCCCCGGAGACCCCGCGGAGTGGGGGCGCGCGCGCGCGCGCGCGCGCGCG 2848
QY 700 TACTAGTC-----ACAGAGCT-----TTAGGGGTG---727
Db 2849 AGCTGCGCTTCCCTACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2908
QY 728 -----GGATT-----CC-----ACTCGTGTGTTT---CTATTTTT 755
Db 2909 GGGAGGACTTCTGATGATTTGAGGTTGACCTTATGCGCTAGGTTTGGTTTGT 2968

```

RESULT 8

```

US-10-137-871-543
; Sequence 543, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-543

```

```

Query Match 48.4%; Score 2790.5; DB 15; Length 3721;
Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;

QY 1 GGGGCC-----CATTC-----GTTTC-----AGCCAG--TCGCCAAG 31
Db 1723 GGGGCGCAGCACCTCAGCTTAGGCTTAGGGACTGCACAGGACTCTCTGGGGCCAGCTCTCCGAG 1782
QY 32 AATCATG-----AAAGTCGCAGTGGCAGC-----56
Db 1783 GACCGCGCGGGGCTGGTGTGCTGTAACCTGCTGTGTAACCTGCTGGTGGCGGCTTCTGTG 1842
QY 57 -----ACCG-----CC-ACCGCGCG-----CGCGGGCC-CC---AGCTGC--86
Db 1843 GTGGAGGCGGTGGTGTGCGGCTTACGCTGGGCTGTGGGCTCGTGGAGCGGCGG 1902
QY 87 GCGCTGAAGCGCGCG-----AAGACA-----GCGAGCGGTGCGGCGGAGGTGG 129

```



```

Db 1783 GACCGCGCGGGCTGGTGGTGAACCTGCTGTGTAACGTCGTGCGTGGCGGCTTCGTG 1842
Qy 57 -----ACCG-----CC-ACCGCGC-----CGCGGC-CC-----AGTCG- 86
Db 1843 GTGGAGCGCTGGTGTCCGGCTTACGCTGGCTGGTGGCTCGGTGAGCGCGG 1902
Qy 87 GCGCTGAAGCGCGC-----AAGACA-----GCGAGCGGTGGCGGCGAGGTGG 129
Db 1903 GAGCTG- -GCCCGCGCAAGGACAGAGAGGCCATCTGCGCGCACGGGCGGCGAGGCGG 1960
Qy 130 TGC- -GCTGTCT- -GTCTGAGC-AGAGCTGGC-CA- -TCTC- - - - - 163
Db 1961 TGTGTAGC- -GTACGCGCTGGCGAGCGAGCGAGCGAGGCTCCCGGGGCGCGGGCGGA 2019
Qy 164 -GCGCTGCGCG- -GCGCGCGCGCGCGCTGCTGCTGCGCTGCTG- - - - -GACGAGC 210
Db 2020 GCGGTGGCGGTGGCGCGCGGGTTCCTCCCGAGGGCGCTGCTGCGCGCTGTATGAGAAC 2079
Qy 211 AGCAGGTAA- - - - -CCTGTGCT- - - - -CTACGATGAAC- - - - -GCGTGTGA 250
Db 2080 GGTGGGCGCAAGCGCACGCTGTGTCAGAGCGGGCGCGCGCGCTGAGCTCGGGGCTGCTG 2139
Qy 251 CTCACGCT- - - - -CA- - - - -AGGAGCTGGTCCCGACCTGCGCCAGAACCGC 293
Db 2140 CCCACGCCGAGCAGACGCGCTGCCGAGAGGCGCTGCCCACTCCGCAACCGCACCC 2199
Qy 294 AAGG- - - - -TGACGAGGTG- - - - -GAGATTCTCCAG- - - - -CACGTATCGACTACA 336
Db 2200 CAGCGCTGGCGCGCGCGCTGGACACCGCGCACCTCCCTGCTCCCGGCTCCGCTTCA 2259
Qy 337 TCAG- -GAGCTTCAGTTGGAGCTGAACTCGAATCGAAG- - - - -TTGG- - - 379
Db 2260 TCTCCCTCTGCTGTGGCGCC- -CGCGCGCGCGCGCGAGCAGCGCGCGCGCTGGGA 2318
Qy 380 - - - - -GACCGCGGGCGG- - - - -AGGGCTG- -CGGTCCGGC- - - - - - - - - 412
Db 2319 GCGGACCGCGCGCGCGCTCTATGCTGCCCGCGCGCGCGCTCCCGCGGACTT 2378
Qy 413 TCGCTCAGCACCT- - - - -CAACGCGAG- - - - -ATCAGCGCTCGACG- - - 453
Db 2379 CCGCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2436
Qy 454 -CGAGG- - - - -CGGATCGTTCCTGCGGAGCATGCGATCTTGTGCTGAAAGCGCTCC 509
Db 2437 CCCTTGACCCAGCT- - - - -CAGCGCGCATGGGCTCCCGCGCGCGCTGAGC- -CGGC 2489
Qy 510 CCC- -AGGG- - - - -GGC- - - - -GGACCCCA- - - - -GCCATCCAG- -G 540
Db 2490 CCGGCGGCGAGCTGAGGAGCGCACTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2549
Qy 541 GGGCA- - - - -AGAGGAATTACGTGCTGTGGGTCTCTGGGTCTCCCGCAACGC- -GCT 585
Db 2550 CCGACCCACACGTTCAACAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2609
Qy 586 CGCC- - - - -GGATCTGA- - - - -GGGAGAACAGACCGATCGG 617
Db 2610 CGCGCGCGCGGCGACAGACTTGGCCACCTCTCCCTATGGGGGGGCG- -GACGAGACTG 2668
Qy 618 CGGCACTGCGCGCTTAA- - - - -TGCAT- - - - -CCAGCTTGGG- - - - - 652
Db 2669 CGCGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2728
Qy 653 - - - - -CTGAG- -GCTGAGGCACTGG- - - - - - - - - - - - - - 670
Db 2729 CAGGAGCGAGAGACGGTCCAGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2788
Qy 671 - - - - -CGAGGAGAGGCGCTCTCTC- - - - - - - - - - - - - - - - - 699
Db 2789 GTCCCCCGCGACCCACCGCGAGTGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2848
Qy 700 TACTAGTC- - - - -ACGAGACT- - - - -TTAGGGGGTG- - - - - - - - - 727

```

```

Db 2849 AGCTCGCCCTCCCTTACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2908
Qy 728 -----GGATT-----CC-----ACTCGTGTGTT-----CTATTTTT 755
Db 2909 GCGAGGACTTTGCTATGATTTGAGGTGACCTTATGCGGTAGGTTTGGTTTTTTTT 2968
Qy 756 GAA-----AAGCAGACATTTTAAAAAATGTC-----CGTT-----TGGT 791
Db 2969 GCAGTTTGGTTTCTTTTGGGTTTCTAACCAATTGCACAACTCGTTCTCGGGTGGC 3028
Qy 792 -----GCTT-----CTCAGAT-----TTCTGAGGAATTG- 816
Db 3029 GGCAGGCGGAGGAGCTTGACCGCGGTGGGAATGGGGGCGCACAGCTGCAGACCTAAG 3088
Qy 817 -----CTTGTATTTATAT-----TACATGATCAGCGAC-----TGAGATAT- 856
Db 3089 CCCTCCCGCCCGCTGAAAGGTCCTCCCGCAACCGCGCGCGCGCGCGCGCGCG 3148
Qy 857 -----TGTTT-----TACAATAGTTCTGTG-----GGCGTGTTTTTTTT 892
Db 3149 CGTGGTGTGCGTCCGCTGTCGTGTGCAAGGGCGCGGAGGTGGGC- -GTGTGTGTGG 3207
Qy 893 ATTAAACAATATTTAGATGGTGAAAAAATAA 926
Db 3208 TGCCAGCGAAGGCTGCTGTGGGCGGTGTGTGCA 3241

```

RESULT 5

```

US-10-141-761-543
; Sequence 543, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-761-543

```

```

Query Match 48.4%; Score 2790.5; DB 14; Length 3721;
Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;
Qy 1 GGGGGC-----CATTC-----GTTTC-----AGCCAG--TCGCCAAG 31
Db 1723 GGGGCGCAGCACTCAGGCTTAGGAGCTGCACAGACTCCTCGCGGCGCAGCTCTCGAG 1782
Qy 32 TATCATG-----AAAGTCGCCAGTGGCAGC----- 56
Db 1783 GACCGCGGGGCGTGTGCGTGAACCTGCTGTGAACGTGTGCGTGGCGGCTTCGTG 1842

```


Db 1723 GGGCCAGACACCTTAGGCTTAGGGGACTGACAGGACTCTCTGGGGCCAGCCTCTCCGAG 1782
QY 32 AATCATG-----AAAGTCGCAGTGGCAGC-----56
Db 1783 GACCGCGGGGCTGGTGTCTGAACTGCTGTGTAACGTCGTGGTGGGGCTTCGCTG 1842
QY 57 -----ACG-----CC-ACGCGCG-----CGGGCC-CC-AGCTGC--86
Db 1843 GTGGAGCGGTGGTGTCTCGGCTTCAGCGTGGGCTGGTTCGTGGGCTTCCTGAGCGCGG 1902
QY 87 GCGCTGAAGCGCGG-----AAGACA-----GGAGCGGTGGCGGCGAGGTGG 129
Db 1903 GAGCTG--GCCCGCGCAGAGACAGAGAGGAGCCATCTCTGGCGCACGCGGCGGGCGAGCGCG 1960
QY 130 TSC---GCTGCTCT--GTCCTGAC-AGAGCGTGC-CA--TCTC-----163
Db 1961 TCTGAGC-GTCAGCGCGCTGGCGAGCGCAGAGCGCAGGGTCCCGGGGCGCGGGCGGA 2019
QY 164 GCGTGTGCGG-GGCGCGGGGCGCGCTGCTCTCCCTGCTG-----GACGAGC 210
Db 2020 GCGGTGTGGGTGGCGCGGGGTTCCTCCCGAGGCGCTGCTGGCGCGCCCTGTATGCAGAAC 2079
QY 211 ASCAGGTAA-----CGTCTCTCT-----CTACGACATGAC-----GCTGTGA 250
Db 2080 GCTGTGGCGAGCGCACCTGCTGTGAGGGGCGGCGCCACGACCTGAGCTCGGGGCTGCTG 2139
QY 251 CTCAGCCT---CA-----AGAGTGTGTGCCACCTGTGCCCAAGAACCGC 293
Db 2140 CCGAGCGCGAGCAGACCGCGCTGCGCGAGAGCGCGCTGCGCACCTCGCACCGCACCGCC 2199
QY 294 AAGG-----TGAGCAGGTG---GAGATTCTCCAG-CAGCTCATGACTACA 336
Db 2200 CAGGCTTGGGCGCGCGCTGGGACACGCGCACCCCTGTCTCCCGGCTCCGCTTCA 2259
QY 337 TCAG-GGACCTTCAGTTCGAGCTGAACTCGGAATCGAAG-----TTGG---379
Db 2260 TCCTCCCTCTGCTGCTGGCGCC-CGCGCGCGCGCGCGAGCAGCGCGCGCGCTGGGA 2318
QY 380 ---GACCCCGCGGGCGG---AGGGCTG-CCGGTCCGGG-----412
Db 2319 GCGGACCGCGCGCGCGCGCTCTATGTGCGCGCGCGCGCGCGCTCCACGCGGACTT 2378
QY 413 TCGCTCAGCACCTT-----CAACGCGAG-----ATCAGCGCGCTGACGG--453
Db 2379 CCGGTCTACCCCGCACCGCGAGCGCGCGCGGGTGTGTGCGCGCGCC--ACGGCG 2436
QY 454 -CGGAGG---GGCATGGTTCCTGCGGAGCATGCTGTGTGCTGCTGAAGCGCTCC 509
Db 2437 CCTTGGACCAAGCT-----CAGCGCGCATGCGCTCCCGGGCGCTGGAG-CGCGC 2489
QY 510 CCC-AGGG--ACC-----GGC-----GACCCCA-----GCCATCCAG-G 540
Db 2490 CCGGACGCGGCGCTGAGGAGGCGCACTGGGCGCGCGCGCGCGCGCGCGCGCGCG 2549
QY 541 GGGCA-----AGAGGATTAAGTGTCTGTGGTCTCTCCCGCGCGCGCGCGCGCG 585
Db 2550 CCGCACCCACAGTTCACAGCGGCGGAGCGCGCGCTGGGACCGCACCGCGCTGCCA 2609
QY 586 CGCC-----GATCTGA-----GGAGAACAGACCGATCGG 617
Db 2610 CCGCGCGCGCGCGCAGACACTTGGCCCACTCTCTCCCTATGGGGGGCG-GGACGAGCTG 2668
QY 618 CCGGCACTGCGCGCTTAAC-----TGCA-----CCAGCTTGGG---652
Db 2669 CGCCCCCGCTAGCGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2728
QY 653 -----CTGAG-GCTGAGGCACTGG-----670
Db 2729 CAGGAGCGAGAGACGGTCCAGAACGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCG 2788
QY 671 -----CGAGGAGAGGCGCTCTCTC-----TGCA-ACC 699

Db 2789 GTCCCCCGCGGACCCACCCCGGAGTGGGGGCGCCCTCCCGCCACAAGAGCAACACC 2848
QY 700 TACTAGTC-----ACCAGAGACT-----TTAGGGGGTG-----727
Db 2849 AGCTCGCCCTCCCTACCCCGGGCGCGAGACGCTGAGACGCTTTGGGGGTGGGTGGG 2908
QY 728 -----GGATT-----CC-----ACTCCTGTGTT---CTATTTTT 755
Db 2909 GGGAGGACTTGTGTAAGATTGAGGTGACCTTATCGCGTAGGTTTGGTTTTTTTT 2968
QY 756 GAA-----AAGCAGACATTTTAAATAAATGGTCA-----CGTT-----TGGT 791
Db 2969 GCAGTTTTTGTGTTCTTTTGGGTTTCTTACCAATTCACACTCCGTTCTCGGGGTGG 3028
QY 792 -----GCTT-----CTCAGAT-----TTCAGAGAAATTC-816
Db 3029 GGCAGGAGGGAGGCTTGACCGCGGTGGGATGGGGGCGCACAGCTGCAGACCTAAG 3088
QY 817 -----CTTTGATTTGATAT----TACAATGATCACCGAC---TGAGAAATAT-856
Db 3089 CCCTCCCGCACCCCTGGAAAGTCCCTCCCAACCCAGCGCGCTGTGTGGGTGTG 3148
QY 857 -----TGTTT---TACAATGATTTCTGTG---GGGCTGTTTTTTTGT 892
Db 3149 CGTGGTGTGGTGGCGCTGTTCTGTGTCAGAGGGCGCGGAGGTGGGC-GTGTGTGTGG 3207
QY 893 ATTAACAAATAATTTAGATGTTGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 926
Db 3208 TGCCAGCGAAGGCTGCTGTGGGCGCTGTGTGTCA 3241

RESULT 3

US-10-146-731-543
; Sequence 543, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Deanoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Zhang, Zemin
; APPLICANT: Zhan, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C323
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-146-731-543

Query Match 48.4%; Score 2790.5; DB 14; Length 3721;
Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;
QY 1 GGGGCG-----CATCT-----GTTTC-----AGCCAG--TCGCCAAG 31
Db 1723 GGGGCGGAGCCTCAGGCTTAGGGGACTGACAGGACTCTCGGGGCGCGAGCTCTCCGAG 1782

QY 1 GGGGCC-----CATTC-----GTTTC-----AGCCAG--TCGCCAAG 31
 Db 1723 GGGGCCAGCACCTCAGGCTTAGGGGACTGCACAGGACTCTCCGCGGCCAGGCTCTCCGAG 1782
 QY 32 AATCATG-----AAAGTCCCAAGTGCACG-----56
 Db 1783 GACCGCGGGGCTGTGTGCTGTAACCTGCTGTAACTGCTGTGCTGTGCTGTGCTGTGCTGTG 1842
 QY 57 -----ACCG-----CC-ACCGCCGC-----CGCGGGCC-CC-----AGTCGC--86
 Db 1843 GTGGAGCGCGTGTGTCCGCGCTTACGCTGTGGGCTGTGTGTGGCTCTCCGTGAGCGGGCG 1902
 QY 87 GCCTGAGCGCGGCG-----AAGACA-----CGAGCGGTGCGGCGGAGGTGG 129
 Db 1903 GAGCTG--GCCCGCGCAGGACACAGAGGCCATCTCTGCGGCGGCGGCGGCGGCGG 1960
 QY 130 TGC-----GCTGTCT-----GTGTGAGC-AGAGGTGCG-CA-----TCTC-----163
 Db 1961 TGTGTAGC-GTACGCGCTTGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2019
 QY 164 GCGCTGCGG-GGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 210
 Db 2020 GCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2079
 QY 211 AGCAGGTAAA-----CGTGTCT-----CTACGACATGAAC-----GCTGTGTA 250
 Db 2080 GCGTGGCCCAAGGCCACGCTGCTGCAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2139
 QY 251 CTACGCGT--CA-----AGGACTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 293
 Db 2140 CCCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2199
 QY 294 AAGG-----TGAGCAAGGTG-----GAGATTCTCCAG--CACGTCAATCGACTACA 336
 Db 2200 CAGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2259
 QY 337 TCAG-GGACCTTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 379
 Db 2260 TCCT 2318
 QY 380 -----GACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 412
 Db 2319 GCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2378
 QY 413 TCCGCTCAGCACCT-----CAACGCGGAG-----ATCAGCGCGCTGACGG--453
 Db 2379 CCGGCTCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2436
 QY 454 -CGGAGG--CGGATGCTTCTGCGGAGCGATGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 509
 Db 2437 CCGTTGACCGAGCT-----CAGCGCGCGATGCGCTTCCGCGGCGCTTGAGC-CGCGC 2489
 QY 510 CCC--AGGG--ACC-----GGACCCCA-----GCCATCCAG-G 540
 Db 2490 CCGGACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2549
 QY 541 GGGCA-----AGAGGAATTACGCTGCTGTGGGTCTCTCCCGCAAGCGC-GCCT 585
 Db 2550 CCGCACCCACAGTTTCAACAGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2609
 QY 586 CGCC-----GGATCTGA-----GGGAGAACAGACGATCGG 617
 Db 2610 CGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2668
 QY 618 CGGCGACGCGCGCTTAC-----TGCA-----CCAGCTGGGG--652
 Db 2669 CGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2728
 QY 653 -----CTGAG-GCTGAGGCACTGG-----670
 Db 2729 CAGGAGCGGAGACGCTGCCAGAACGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2788

QY 671 -----CGAGGAGAGGGCGGCTCTCTC-----TCAC-ACC 699
 Db 2789 GTCCCCCGCGGACCCCGCGGAGTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2848
 QY 700 TACTAGTC-----ACGAGAGCT-----TTAGGGGTG--727
 Db 2849 AGCTCGCCCTCCCGCTTACCGGGGCGGAGGACGCTGAGACGGTTTGGGGGTGGGTGGGC 2908
 QY 728 -----GGATT-----CC-----ACTCGTGTGTTT-----CTATTGTTT 755
 Db 2909 GGGAGGACTTGTGATGATTTGAGGTGACCTTATGCGGCTAGGTTTGTGTTTTTTTTT 2968
 QY 756 GAA-----AAGCAGACATTTTAAAAATGGTCA-----CGTT-----TGGT 791
 Db 2969 GCAGTTTGTGTTTCTTTGCGGTTTCTTAAACCAATTTGCAACTCCGTTCTCGGGGTGGC 3028
 QY 792 -----GCTT-----CTCAGAT-----TTCTGAGGAATG-816
 Db 3029 GGCAGGAGGGGAGCTTTGGACCGCGGTGGGGAATGGGGGCGGCGGCGGCGGCGGCGGCGG 3088
 QY 817 -----CTTGTATTTATAT-----TCAATGATCAGCGAC-----TGAGATAT-856
 Db 3089 CCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3148
 QY 857 -----TGTTT-----TACAATAGTTCTGTG-----GGGCTGTTTTTTTTT 892
 Db 3149 CGTGGGTGGGTGCGGCTGTTCTGTCAGAGGGCGGCGGAGGTGGGC-GTGTGTGTGGC 3207
 QY 893 ATTAACAATAATTTAGATGTTGTAATAAAAAA 926
 Db 3208 TGCCAGGAGGCTGCTGTGGCGGTGTGTGTCAA 3241

RESULT 2

US-10-123-155-543
 ; Sequence 543, Application US/10123155
 ; Publication No. US2003006794A1

GENERAL INFORMATION:

APPLICANT: Baker Kevin P.
 APPLICANT: Beresini, Maureen
 APPLICANT: DeForge, Laura
 APPLICANT: Desnoyers, Luc
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Sherwood, Steven
 APPLICANT: Smith, Victoria
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Watanabe, Colin K
 APPLICANT: Wood, William
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 TITLE OF INVENTION: ACIDS ENCODING THE SAME
 FILE REFERENCE: P3330R1C30
 CURRENT APPLICATION NUMBER: US/10123,155
 CURRENT FILING DATE: 2002-04-15
 Prior Application removed - See Palm or File Wrapper
 NUMBER OF SEQ ID NOS: 550
 SEQ ID NO 543
 LENGTH: 3721
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-10-123-155-543

Query Match 48.4%; Score 2790.5; DB 14; Length 3721;

Best Local Similarity 39.7%; Pred. No. 1.3e-131;

Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;

QY 1 GGGGGC-----CATTC-----GTTTC-----AGCCAG--TCGCCAAG 31

Db 631 TGTTCCTCAATGACTCTCAACTGTAAACTGGAATTTGAAGCTATGCG--TATTCGAC 688
Qy 319 AGCAGCTCATCGACTACATCAGG-GACCTTCAGTTGGAGCTGAATCGGAATCCGAAGTT 377
Db 689 GGC-CGAAATCGAGTACAAATGCTGACGTC-GAAGGAGCCGAATTTGTCAGACGCGTC 746
Qy 378 GGGACCCCGGGGGCGAGGCTGCCG-----GTC-----CGG 410
Db 747 AGGCCGAGCGGAACATCGAATGTCGAGTTATTAATTCATAAATCTGCCAAAACGG 806
Qy 411 GCTCCGCTCAGCAC-----CC-----TCAACGG----- 433
Db 807 ACATTTGCCAGCACTTCATCGGGACCTACTCTCGTCTACGGTTAGTTTTCATATTGAT 866
Qy 434 CGAGATCAGCG-----CCCTGACGGCC----- 455
Db 867 CGCGA-CAGCGGCTTCTACTTTCTCAAAATATTTTCCCTGCGAGCTCTGCTGATGTTT 925
Qy 456 -----GA-----GGCGGAT-GCGTTCC-----TGC 475
Db 926 ATCATGGATCTCATCTGGATCAATCGTGACTCGGCGCTTCGGGAACCTAATCGGTAC 985
Qy 476 G--GACGAT-----CGCATCTTGTGCTGCTGAAGCG-----CCTCCCCAG 514
Db 986 GATGACGGTCTCACTGAGACTCATCTATGAC-CGGNAACCAATCGACGCTTCCACCAG 1044
Qy 515 GGACCGGCG-----GAC-----CC-----CAGCCATC-CAGG-----GGG 543
Db 1045 TTGCTATGTAAGCGGTTGATGATTCTCGGTTTCTGCTATCTTCTGGTTATACG 1104
Qy 544 C-AAGAGGAATAGTCTCTGTG-GGTCTCCCC-----AACGGCCCTCGCCGA 592
Db 1105 CGTTGATCGAGTAGC--CCTGTGTGCTTACTCAAAAAGAGAACGAGGATCGTCGA 1161
Qy 593 TCTGAGGGAGACAGACCGATCGGCGCC-ACTGGGCC-----CTTAACATG 638
Db 1162 --GAAGAGAGAGAGCGAGGATAAACTGTTCGCGGACACCTGATATCTTCACGA 1219
Qy 639 CATCCAGCTGGGCTGAGGCTGAGGCACTGGGAGAGAGGCGC--TCTCTCTGCAC 696
Db 1220 CGTCC-SCCT--TGCCGA-----ATGCACATGCAAGC--GGCTCCAACTCGATCATC 1268
Qy 697 ACC-----TACT-----AGTCAC-----CAGAGACTT--TAGGG 723
Db 1269 GCGCTCATCAAGCAGTCGAATCGATTCTGTGACGTACAGTCACATGACATCGTCAGC 1328
Qy 724 GGT---GGGATTCACCTCGT-----GTGTT---TCTATTTTTC----- 756
Db 1329 CGTCCCGGTTCTCTTGTGTTTCATCTTGTCAACACTCTTCTGGCTGATTCCTACTG 1388
Qy 757 ---AAAAGCAGACATTT-----TTAAAAATGGTCAG-----TTTGGTGTTC- 796
Db 1389 TACAAATCCAGCGTCTGCCGTATATTAGTGAACACGAGGGTACCGGTCGGATGCTCCA 1448
Qy 797 -----TCA--GATTTCTGAGGAATGCTTTGTATTGTATATTAATCAATGATCAC----- 843
Db 1449 GACCTTCATTAATCTCAATCCAACCTCTCATCATTTTCCATTTCGAATATCTCTTTTC 1508
Qy 844 -----CGACTGAGAAATATCTTTTACAAATAGTTCTGTGGGCTGTTTTTTGTTAT 894
Db 1509 TTGCACAGAGCCCTTTTCGTTTTTTTTTTTATGATTTATTTTACGGATTTTAGATAA 1568
Qy 895 TAAACAAA-----TAATTTAGATCGTGAAAAAATAA 926
Db 1569 TGCACAGATGCTCATTCCTCAATAAATTTATTTTAAATGTCGAAAAAATAA 1622

Search completed: May 7, 2004, 15:20:45
Job time : 43 secs

Matches 541; Conservative 0; Mismatches 354; Indels 448; Gaps 77;
Qy 7 CATTCTGTTTCAGC---CAGT---CGCCAGAAATCATGAAGTCGCCAGTGGCAGACCC 59
Db 63 CAAGCTGGGTGGCTTTTCATCAACACCAAGAACAGCTAAGCAGCTGCTCCAGGACCC 122
Qy 60 ----GCCACCGCGCGCGCGCCAGCTGCGCG---CTGAAGCGCGCAAGACAGCGAG 112
Db 123 CTGCTCCACTGCAGAAAAGGGTCCACTTAC-CATTTCTGCA-CAGGAAAATGTAAG 179
Qy 113 CCGTG---CGGGAGGTTGGTGGCTGTCTGTGAGCAGAG-CGTGGCCATCTCG 164
Db 180 TGGTGTATTACCGGGAC--TGATCCCTTTGAATC---CAGAAGCCATGATGAATCA 233
Qy 165 CG---CTGCCCGGGGCG---CGGGGCG---CGCTGCGCTG---CCCTGCTGGA 205
Db 234 CTATCCAGCCAGGAGACATGATCATGTGGATGAAGCCAACTGGAGAACCGGCTGGC 293
Qy 206 C--GAGCAGCA---GGTAAC---GTGCTGTCTACGACATGAACGG--- 244
Db 294 TTGGAGGAGAAATTAAGGAAAGAGACAGGGTGGTTCCCTGCAAACTATGACAGAAATCC 353
Qy 245 CTGTATC-----TCACGCTCA-AGGAGCTGGTGC-----CA-----CCCTGCC 284
Db 354 CAGAAATGAGTTCCCGTCCAGTGAACCACTGACTGATTCACATCTGCCCTGCCC 413
Qy 285 CAGAACCG--CAAG-GTAG---CAAGGTGAGAT-----TCTCAG-----CA 322
Db 414 CCAAACTGGCTTGGGTGAGACCCCGCCCTTTGGCAGTAACCTCTTCAGAGCCCTCA 473
Qy 323 CGTCATC-----GACTACATCAG-----GACCTTCAGT-----TGGAGC 357
Db 474 GACCCCTAATTAACCTGGCGGAGTTTCACTCCAGTGGCCACACGACGATGAGAAC 533
Qy 358 TGAATCGGAATCC---GAAG-TTGGACCCCGG---GGCGCGAGGGT- 400
Db 534 CAGAAACGATAACTGGGATGATGGGCGACGCCAGCCCTCTCTACCGTTCCAAGTGGCG 593
Qy 401 GCGGTCCGGC---TCGC-CTCAGACCTCAACGGCGA---GATCAGCGCCCTGAC 451
Db 594 GCAGTTAAGGAGAGGTTCGCTTTACTTCAGACACGCGCACTGCTC-CTCCCGCTC 651
Qy 452 -----GGCC-----GAGGGCGCATGC-----GTTCTCT-- 473
Db 652 TCCTGTGCTAGGCGAGGTGAAAAGGTGGAGGGGCTACAAGCTCAAGCCCTATATCTTG 711
Qy 474 ----GC-----GGAGATCGCATCTGTGCTGAAGCG---CCTCCCC--CAGGGA 517
Db 712 GAGAGCAAAAAGACAACCACTTAAATTTTAACAAAAATGATGATCACCCTCTGGA 771
Qy 518 CCGGC-----GGA-----CCCCAG--- 531
Db 772 ACAGCAAGCATGTGTGTTTGGAGAGTTCAAGGTCAAGGGTGTGTTCCCCAAGTC 831
Qy 532 -----CCATCCAGGGGGCAAGAGGA--TTAGTGC-----TC 562
Db 832 TTACGTGAACACTCATTCAGGGGCCATAAGGAAGTCTACAGCATGGAATCTGTGTTCTC 891
Qy 563 TGTEGG-----TCTCCC-----CCAA--CGCGCTCCCGCG 591
Db 892 AGAGAGTCTCTAGTCTAAAGCGAGTAGCTCTCCAGCAGCAAGCCGAGCAAGCGG 951
Qy 592 ----ATCTGAGGAGAACAGACCG-----ATCGGGGGCACTGCGCCCTTAA 635
Db 952 TCGTTTTCGGAGAGAAATTTGCCACAGGTTATTGCTCTATACCCGCCACCG-GCCCGCAG 1010
Qy 636 CTCG---ATCCAG-CCTGG--GGCTGAGCTGA-----GGCACTGGCG----- 672
Db 1011 CAGTCACTCTGCCCTGTGAGCTGATTTTGTATCCGAAAAGAACCCAGGTGATGG 1070
Qy 673 -AGGAGAGCGGCTCT--CTC-TG-----CACACCT-----ACTAGTCAACAGAGACT 717
Db 1071 TGGGAGAGAGAGCTGCAAGCACCTGGGAAAAAGCCGACAGATAGTGGTTCCAGCTAAT 1130

Qy 718 TTAG-----GGGTGGG--ATTCCACT-CGTGTGTTCTATTTTTGA 757
Db 1131 TATGTAAGCTTCTAAGCCCTGGACGAGCAAAATCACTCCACAGAGCCACCTAAGTCA 1190
Qy 758 AAGCA-----GACATTTTAAATAATGGTCACTGTTGGTGTCTCT-----CAGATTT 804
Db 1191 ACAGCATTAGCGCAGTGTCCAGGTGATTTGGGAT--GTACGACTACACCGCGCAGAATG 1248
Qy 805 CTGAGGAAATTTGCTT-----GTATTGTATATTACAATGA--TCACCG 845
Db 1249 ACATGAGCTGGCTTCAACAAGGGCCAGATCATCAAGCTCTCAACAAGGAGCCCTG 1308
Qy 846 ACTG--AGAAATATGTTTACAATAGTCTGTGGGCTGTTTTTTTGTATTAAACAAT 903
Db 1309 ACTGGTGAAGAGAGAGT-CAATGGACAAGTGGGGCTCTT-----CCCATCAAT 1358
Qy 904 AATTAGATCGTGAATAAAAAA 926
Db 1359 TATGT-GAAGCTGACACAGACA 1380

RESULT 14

US-09-627-650B-1
; Sequence 1, Application US/09627650B
; Patent No. 640872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1652
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-1

Query Match 40.4%; Score 2327; DB 4; Length 1652;
Best Local Similarity 37.3%; Pred. No. 8,9e-98;
Matches 579; Conservative 0; Mismatches 321; Indels 554; Gaps 84;

Qy 1 GGGGC-CCATTC-----TGTTTCAGCCAGTCGC-----CAAGAATCA 36
Db 95 GAAGCTCCATTCGAGAAATGGCTCGTCCATTACACTTATCGTACTCTCTCCGCACA 154
Qy 37 TGAAGTCGCCA-GTGG-----CAG-----CACCGC-CACGCGCG 69
Db 155 TCTGTGCTACATGTTGTTGTGACAGAGATGAGGACTACATATCAACTCACTCT 214
Qy 70 C-----CGCG-----GGCCCCAGTGC CGCTG-----AAGGCC--G 99
Db 215 CTATCATGTTCTCGATAGACTCAGAAATCGCACTTATGATAAAAGATTACGCCAG 274
Qy 100 GCAAGACAGCGCGGTGGCGGCGAGGTGG-----TGCGCTG-----TCTGTCTGAG 146
Db 275 GTATGGTGAAGC-----CAGTCGACGTGGAATTACGATACAGCTTCTTCACTCTG 330
Qy 147 CAG-----AGC-----GTGGCCATCTC-----GCG-CTGCCGGGCG-- 176
Db 331 CAGTTTCAGAAGTTGATATGAGACTTTCACATTAGACTTCTATACATCGCTCAAAACGTGGCAAG 390
Qy 177 -----GCC-----GGGG-----GGCCTGCTGCTGC 196
Db 391 ACCCTGACTAGCTTCGGAAGTCTTGAATTGGAGCTTTCAAAAGAAATGACTCACTTA 450

QY 59 CGCC--ACGCGCGCGCGCGCC--CCAG--CTGCGG--CTGAAGCGCGCAA 103
DB 1194 CGCGCAACACAGCAGCGAAACGAGATGCCAATGTTCAACGCGAGCGCGAGCGCGCAA 1253
QY 104 -----GACA-----CGAG--CGGTGGCGGCGAGTG--GT 130
DB 1254 TAATAATTCATAGAAATGACATCTATGTCGAAAATTCGACGCTGCCAAAAGCTATGT 1313
QY 131 GC--GCTGTG--TGT--CTGAGCAGAGC--GTGGCCATCTC 163
DB 1314 ACAGGCTGACTGTACTTTGCGCGGACACAAATTCCTATGATCCATGATGAGATGCC 1373
QY 164 -----GGCG--TGCGG-----GGCGCGG--GG 183
DB 1374 AGAAAATGTTGATGTCGCGAGCAATCCAAATGATGCAACATCCAGCTGTTGTACAGAGCG 1433
QY 184 CGCGCTCG--CTGCCCTGCT--GGACGAGCAGCAGGTAA--ACGTGCTG 227
DB 1434 CGACATACCTATGCGCGCTCCATTCGCGCGCGGAAAAGGCTTCCAGACATGCTG 1493
QY 228 CTCTACGACATGAACGGCTG-----GTG-----247
DB 1494 C-CAACG--ATGGACGCGCTCAAAAATCGAATAAGCTTAGCCGATACGGTTTCCCATGTC 1550
QY 248 TTACTACGCTCAAG--AGCTGGT-----GC-----CCAC-----277
DB 1551 TTCTCTATCTTCAATATAGTCTACTGTTGTATGAAATATCTAAGCTTAACTCGTC 1610
QY 278 ---CCTGCCCGAG--AACCGAAGTG--AGCA-----AGTGAGATTTCTCCAGCAC 323
DB 1611 GGACAAGATCCAGGAGAACCAAGTGGCAGCAGATCCACTGATGGCTATTCGACGCG-C 1669
QY 324 GTCATCGACTACATCAGG-CACCTTCAGTTGAGCTGAACTCGGAATCCGAACTGGGAC 382
DB 1670 GAAATCAGTACAAATGGTGTAGCTC-GAAGGAGCGGAATGTTCCAGCAGGTCAAGC 1728
QY 383 CCCCGGCGCGGCGGCTGCCG-----GTC-----CGGCTCC 415
DB 1729 CGACGGAACATCGAACTGTCGAGTTATAAATCACTAAATTCGCAAAAACGACACT 1788
QY 416 GCTCAGAC-----CC-----TCAACG-----CGAGA 438
DB 1789 TGCCAGCAGCTTCATCGGGAGCCTACTCTGCTACGGGTAGTTTCATATTCATCGCA 1848
QY 439 TCAGCG-----CCTGACGCGC-----455
DB 1849 -CAGCGGCTTCTACTTCTTCAATATTTTCCCTGCGCAGCTCGTGTATTTATCAT 1907
QY 456 -----GA-----GGGCGAT--CGGTTCC-----TGCG--GA 478
DB 1908 GGATCTCATTCGATCAATCGTACGCGCGCTTCGGAACCCCTAATCGGTACGATGA 1967
QY 479 CGAT-----CGCATCTTGTGCTGTAAGCG-----CCTCCCGCAGGAGCC 519
DB 1968 CGGTGCTCACTGAGACTCATCTTATGAC-CGGAACCAATCGAGCTCTCCACAGTTGCC 2026
QY 520 GGCG-----GAC-----CAGCCATC--CAGG-----GGGC--AAG 547
DB 2027 TATGTAAAAGCGTTGATGTATTCCTCGGTTTCTGCTATCTTCTGTTATCTGCGGTTG 2086
QY 548 AGGAATTACGCTGCTGTG--GGTCTCCCG-----AACGCGCTCGCGGATCTGA 597
DB 2087 ATCGATAG--CCTGTTGCTCTACTCAAAAAGAGAGCAGGATGTCGGA--GAA 2141
QY 598 GGGAGAACAGACCGATCGCGGCG--ACTGGCGC-----CTTAAGTGCATCC 643
DB 2142 GAGAGAGAGACGAGCAGTAAACCTGCTCGCGCAGCAGCTGATATTTCTCAGAGCTCC 2201
QY 644 AGCTGGGCTGAGGCTGAGGCTGAGGAGGAGGCGG--TCCTCTCTGCACACC--699
DB 2202 -GCCT--TGCGGA-----ATGCAATGCAAGC--GGCTCCACCTGATCATCGCGT 2250
QY 700 -----TACT-----AGTCAC-----CAGAGACTTT--TAGGGGT--726

RESULT 13
US-08-630-915A-37
; Sequence 37, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds LLP
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/630,915A
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-630-915A-37
Query Match 40.5%; Score 2337.5; DB 4; Length 1400;
Best Local Similarity 40.3%; Pred. No. 2.7e-98;

ORGANISM: Caenorhabditis elegans
US-09-627-650B-9

Query Match 41.3%; Score 2383.5; DB 4; Length 2601;
Best Local Similarity 38.0%; Pred. No. 3.6e-100;
Matches 589; Conservative 0; Mismatches 313; Indels 647; Gaps 87;

Db Qy 2027 TATGTAAGACCGTTGATGTTATCTCGGTTCTGCTATCTCTGTTTACTGCGGTG 2086
548 AGGAATTACGTGCTCTGTG-GGTCTCCGCC-----AACGGCCTCGCGGATCTGA 597
2087 ATCAGATAGC---CCTGTGCTCTACTCAAAAAGAACGAGGATCGTCGA--GAA 2141
598 GGGAGACAAACACGATCGCGGCC-ACTGGGCC-----CTTAATCGCATCC 643
2142 GAGAGAAAGACGAGCATAAACCTGCTCGCGGACACCTGATATTTCTTCACGCGTCC 2201
644 AGCCTGGGGCTGAGGCTGAGGACTGCGAGGAGGAGGCGC--TCCTCTCTGCACACC-- 699
2202 -GCCT--TGCGCA-----ATGCACATGCAACGC--GGCTCCAACCTCGATCGCGCT 2250
700 -----TACT-----AGTCAC-----CAGACACTT--TAGGGGT-- 726
2251 CATCAAGCAGTCGAATCGATTCTGTGTCAGTCAGTCACATTCACATCGTCAGCGCTGC 2310
727 -GGGATTCCTACTCGT-----GNGTT---TCATTTTTTG-----AA 758
2311 CGGTTTCTCTGTTTTCATCTTGTTCACACATCTCTCTGCGCTGATCTACTGTACAA 2370
759 AAGCAGACATT-----TTAAAAATGGTCAG-----TTTGGTGTCTC-- 796
2371 ATCCAAGCGTCTGCGGTATATTAGTGAACACGAGGCTGACCGTTGCGATGCTCCAGCCT 2430
797 TCA-GATTTCTGAGGAATGCTTGTATGTATATATACATGATCAC----- 843
2431 TCATTAACTCAATCCACCTTCTCATCTTTCCATTTCCATTCGATATCTTTTTCTTGCA 2490
844 ---CGACTGAGAATATTGTTTACAATAGTCTCTGCGGCTGTTTTTTTCTTATTAAAC 899
2491 CAGAAGCCTTTTTTCGTTTTTTTTTATTGATTTATTTTACGGATTTTATAGATATGCAC 2550
900 AAA-----TAATTTAGATGTGAAAAA 926
2551 AGATGCTCTCATTTGCTCAATAAATTTATTTTAAATTAATAAAAAA 2599

RESULT 12
US-09-436-063C-9
; Sequence 9, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-9

Query Match 41.3%; Score 2383.5; DB 4; Length 2601;
Best Local Similarity 38.0%; Pred. No. 3.6e-100;
Matches 589; Conservative 0; Mismatches 313; Indels 647; Gaps 87;

Qy 2 GGGCC-----CATTCTGTTTC-----AGCC-----AGT- 24
Db 1074 GGGCTGGATGTGTTCTTAATTTTGTTCGTAATGTTATCGCTCGTTCGTCGAGTA 1133
Qy 25 GCGCA-AGAAAT-----CATGAA-AGTCG-----CCAGTGGGAGCA-----C 58
Db 1134 CGCCATAGTATCTTACATGAATAAACGACTGTCTCTCCGCGGAAAAACGAGAAAGC 1193

Db Qy 1074 GGGCTGGATGTGTTCTTAATTTTGTTCGTAATGTTATTCGCTCGTTCGTCGAGTA 1133
25 GCGCA-AGAAAT-----CAGTAA-AGTCG-----CAGTGGCAGCA-----C 58
1134 CGCCATAGTATCTTACATGAATAAACGACTGTCTCCGACGGGAAAAACGAGAAAGC 1193
59 CGCC--ACCGCGCGCGCGGCC-----CCAG-----CTGCGCG-----CTGAAGCGCGGCAA 103
1194 GCGGACACACAGCAGCGAAACGAGATGCCAATGTTCAACGCGAGCGCGGAGCGCGCAA 1253
104 -----GACA-----GCAG--CGGTGCGGCGAGGTG-----GT 130
1254 TAATAATTATACGAATGACATTTATGTGCGAAAATTCGACCGTGCCTGCAAGCTATGT 1313
131 GC--GCTGTC-TGT-CTGACGAGC-----GTGCGCATCTC 163
1314 ACAGGTGACTTGTACTTTGCGGACACAAATCTCTATGATCCATGATGAGATCCC 1373
164 -----GGCG-TGCGG-----GGCGCGG-GG 183
1374 AGAAATGTTGATTCGCGGAGATTCCTCAATGATGCAACATCCACGCTTGTCCAGACGG 1433
184 CGCGCTGC--CTGCGCTGCT--GGAGGAGAGCAGGTAA-----ACGTGCTG 227
1434 CGCATACGCTATGCGCGGCTCAATTCGCGCGCGCGGAAAGGCTTCCAGACATGCTG 1493
228 CTCTACGACATGAACGCGCTG----- 247
1494 C-CAACG--ATGGAGCGCTGCAAAATCGATAAGCTTAGCCGATACGGTTTCCATTGTC 1550
248 TTACTACGCTCAAGG-----AGCTGCT-----GC-----CCAC--- 277
1551 TTTCTCTATCTTCAATATAGTCTACTGTTGTTATATGAATAATCTAAGCTTAACTCTGTC 1610
278 ---CCTGCCCGAG--AACCGCAAGGTG--AGCA-----AGGTGGAGATTTCTCCAGCAC 323
1611 GGACAAGATCCAGGAGAACGACAGTGGCAGCAGATCCACTGATGCGTATTCGACGCG-C 1669
324 GTCATCGACTACATCAGG-GACCTTCAGTTGGAGTGAATCGGAATCCGAATTTGGGAC 382
1670 GAAATCGAGTACAAATGCTGTACGTC-GAAGGAGCGGAATTTGTCACACGCGTCAAGC 1728
383 CCCCGGGCGGCGAGGCTGCGG-----GTC-----CGGCTCC 415
1729 CGACCGAACATCGAATGTGCGATTATAAATTCATTAATCTGCCAAAACGAGACT 1788
416 GCTCAGCAC-----CC-----TCAACGG-----CGAGA 438
1789 TGCCAGCACTTCATCGGGGACCTACTCTGCTACGCGTTAGTTTATATTTGATCGCGA 1848
439 TCAGCG-----CCTGACGCGC----- 455
1849 -CAGCGGCTTCTACTTTCTCAATAATTTTCCCTGCGAGCCTCGTGTGTTTATCAT 1907
456 -----GA-----GCGCGCAT-GCGTTCC-----TGCG--GA 478
1908 GGATCTCATCTGATCAATCGTACTCGCGCGCTTCGGAACCTTAATCGGTACGATGA 1967
479 CGAT-----CGCATCTTGTGCTGAGCG-----CCTCCCGCAGGAGC 519
1968 CGGTGCTCCTGAGACTCATCTTATGAC-CGGAACCAATCGAGCTTTCACAGTTGCC 2026
520 GGCG-----GAC-----CC-----CAGCCATC--CAGG-----GGGC-AAG 547

; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 2088
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-551-853D-13

Query March 41.4%; Score 2385; DB 4; Length 2088;

Best Local Similarity 37.6%; Pred. No. 2.7e-100;

Matches 600; Conservative 0; Mismatches 312; Indels 682; Gaps 93;

```
QY 1 GGGGCCCATCTCT-GTTTCAG-CCAGT-----CGCCAGA-ATCATGAAG-TCG 45
DB 373 GATGCCC-TCTCGTCTCTGACAGGTGMAATCTTACACAGAGAGGATGGATGTTG 431
QY 46 CCAGTGGCAGACCCG-----CACCGCGCGCGCGCGCGCGCTGCGCGTGAAGGCCG- 99
DB 432 CGAACTCATCTCTCACTGGCAGACACCGTCCGCAAGAGAGATGTCAGTGAGAGAGTACCA 491
QY 100 ---GCAGACAGCG-----AGCGTGGCGG-----CGAGTG-----128
DB 492 CTTGCTATGACTACGGCATGTTGCTCCCTGCGGAATTCACAAGTTCGAGGGGTAGAGTT 551
QY 129 ---GTG-----CGCTCTCTG-----TCTG-----AGCAGAGCGTGG 156
DB 552 TGTGTGTTGCCACTGCTGCTGAAAGAAAGTACAAATGTGATCTGTGATGCGGAG-GAGG 610
QY 157 CCATCTCGCGCTG-CGCGCGCGCGCGCGCGC-CTGCGTCCCTGCTGGAGC-----207
DB 611 ATGACTCG-GATGCTGTGTGGCGGAGCAGACAGACTATGAGATGGAGTGAAGAC 669
QY 208 -----AGCAGC--AGTAAACGT-GTGT-----CTCT 231
DB 670 AAGTAGTAGAAGTAGCAGAGGAGGAAGTGGTGGAGTGAAGAGAGAGAGAGCGCAT 729
QY 232 -ACGACATGAAC-----GCTGTGTTACTCAC-----GGCTGTTACTCAC-----255
DB 730 GATGACGAGGACGATGAGTGTGATGAGTGTAGAGGAGAGGCTGAGAACCTACGAA 789
QY 256 ---GCTCTAAGGAG-----CTGG---TGCCCACTCTGCG-CCAGAACCGCA-AG---GTG 299
DB 790 GAAGCCACAGAGAGAACCAACAGCATTTGCCACCAACCAACCAACCAACCAACAGTCTGTG 849
QY 300 AGCAAGCTG---GAGATTCT-----CCAGCAC-----GTCACTGAC---TAC 335
DB 850 GAAGAGGTGTTGAGTTCTTACACAGCAGCCATACCCCTGATGCGTTGACAGTAT 909
QY 336 ATC-AGGACCTTTCAGTTGGAGCTGAAC-----TC--GGAATCCGAAG-----TT 377
DB 910 CTCGAGACACCTGGGGATGAGAAATGAACATGCCCATTTTCCAGAAAGCCAAAGAGAGCCT 969
QY 378 GGGACC-----CCCGGG-----389
DB 970 GAGGCCAAGCACCAGAGAGAAATGTCACAGGTTCATGAGAAATGGAGAGGCGCAGACGT 1029
QY 390 -----GGCCGAGGGGTG-----CCGG-----TCCGG-----411
DB 1030 CAAGCAAGAACTTGCTTAAGCTGATAAGAGGAGGAGTATCCAGCATTTCCAGGAGAA 1089
QY 412 -----CT-----CCCG-----TCAGC-----ACCTCA--429
DB 1090 GTGGAATCTTTGGAACAGGACAGCAACCAACGAGAGACAGAGCTGTTGGAGACACATG 1149
QY 430 -----ACGCCAGATCAG-----CGCC-CTAC-GGGCGAGG-----459
DB 1150 GCAGAGTGAAGCCATGCTCAATGACCGCGCGCGCTGGCCCTGGAGAACTACATCACC 1209
QY 460 ---CGGCATGC-GTTCCTCGGACGATGCATCTTGTGT-----CGCTGAA-----501
```

```
DB 1210 GCTCTGAGGCTGTCTCTCC--TCGGCTCGTCACTGTTCATATGCTAAAGAGATG 1267
QY 502 ---GGGC-----CTCCCCCAGGAGC-----GG 521
DB 1268 TCCGGCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1327
QY 522 CGGACCCC-----AGCC-----ATCCAGGG-----GGCA-----AGAG-GAATT 554
DB 1328 TGGATCCCAAGAAAGCCGCTCAGATCCGCTCCAGGTTATGACACACCTCCGCTGATTT 1387
QY 555 ACCTGC-----TCTGTGGGTCTCC-----CCACGCGCCT-C-----GCCG-----590
DB 1388 ATGAGCGCATGAATCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1447
QY 591 -----GATCTG-----AGGAGAGAA-----CA 609
DB 1448 TTCAGGATGAAGTTGATGAGCTGCTTCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1507
QY 610 CC-----GATCGG-----CGGCCAC--TGCGC-----629
DB 1508 CCAACATGATTAGTGAACCAAGAGATCAGTTACGGAAACGATGCTCTCATGCCATCTTGA 1567
QY 630 CTTTAACCTGCATCCAGCTGGGCT-----GAGGCTGAGG---CA-CTGG-----670
DB 1568 CGAAACGAAACCAACCGTGGAGCTCCTCCCGTGAATGAGAGTTGAGCTTGGACATC 1627
QY 671 -----CGAGAG-----AGGGGCT--CCTCTCTGCACACCTACT-AGTCACC-----AG 712
DB 1628 TCCAGCCCTGGCATTTCTTTGGGCTGACTCTGTGCGAGCAACACAGAAACGAAAGTTG 1687
QY 713 AGACTTTAGGG-----GTGGGAT-TCCACTCG-----TGCT 744
DB 1688 AGCGTTGTATGTCGCGCTGCTGCGCAGCAGACTGACCATCAGCAGGTTCTGGT 1747
QY 745 TCTATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 800
DB 1748 TGACAAATATCAAGACGAGAGAGAGATCTCTGAAGTG---AAGATGGATGAGAAATCCGAC 1804
QY 801 ATTTCTGAGGAAAT--TGCTTTGTTATTTATTTACATGATCAGCAGCTGAGAAATTTG 858
DB 1805 ATGACTCAGATATGAAGTTCATCAATAAAAT--GGTGTCTTTG-CAGAGAGATG 1861
QY 859 TTT--TACAAATGTTCT-----TGTTGGGCTGTTTTTT-----TGT 891
DB 1862 GTTCAACCAAGGTGCAATCATTTGGACTCATGTTGGCGGTGTTGTATAGCAGATGA 1921
QY 892 TATTAACCAATAATTTAGATGTTGTAATAAAAAA 925
DB 1922 TCTTCATCACCTTGT--GATGCTGAAGAGAA 1953
```

RESULT 11

US-09-627-650B-9

; Sequence 9, Application US/09627650B

; Patent No. 6406872

; GENERAL INFORMATION:

; APPLICANT: Bamber, Bruce

; APPLICANT: Jorgensen, Erik

; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and

; TITLE OF INVENTION: Methods Related Thereto

; FILE REFERENCE: 21101.0009U3

; CURRENT APPLICATION NUMBER: US/09/627,650B

; CURRENT FILING DATE: 2000-07-28

; PRIOR APPLICATION NUMBER: 09/436,063

; PRIOR FILING DATE: 1999-11-08

; PRIOR APPLICATION NUMBER: 60/107,727

; PRIOR FILING DATE: 1998-11-09

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 9

; LENGTH: 2601

; TYPE: PRN

;; CURRENT APPLICATION NUMBER: US/09/548,367D
;; CURRENT FILING DATE: 2000-04-12
;; PRIOR APPLICATION NUMBER: US 60/155,493
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: US 09/404,133
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: PCT/US99/20881
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: US 60/101,594
;; PRIOR FILING DATE: 1999-09-24
;; NUMBER OF SEQ ID NOS: 73
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 13
;; LENGTH: 2088
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; US-09-548-367D-13

Query Match 41.4%; Score 2385; DB 4; Length 2088;
Best Local Similarity 37.6%; Pred. No. 2.7e-100;
Matches 600; Conservative 0; Mismatches 312; Indels 682; Gaps 93;

QY 1 GGGGCCCATCT-GTTTCAG-CCAGT-----CGCCAGA-ATCATGAAG-TCG 45
Db 373 GATGCC-ITCTCGTCTCAGAAAGTGCAAAATTTACACAGAGAGATGGATGTTG 431

QY 46 CCAAGTGGCAGCACCGC-----CACCGCGCGCGGCCCGAGCTGCGCTGAAGCGCG- 99
Db 432 CGAACTACTCTTCACTGGCACACCGCTGCGCAAGAGACATGCACTGAGAGAGTACCAA 491

QY 100 ---GCAAGACAGG-----AGCGGTGCGG-----CGAGGTG----- 128
Db 492 CTTGCATGACTACGGCATGTTGTGCGCTCGGAAATTGACAACTCCGAGGGGTGAGTT 551

QY 129 ---GTG---CGTGTCTG-----TCTG-----AGCAGAGCGTG 156
Db 552 TGTGTGTGCTCCACTGCTGAAGAAAGTGAATGTTGTTCTGCTGATGCGGAG-GAGG 610

QY 157 CCAATTCGCGCTG-CGCGGCGCGCGGCGCGC-CTGCGTCCCTCTCTGAGC----- 207
Db 611 ATGACTCG-GATGTCTGTGTGGCGGAGCAGACACTATGACAGTGGAGTGAAGAC 669

QY 208 -----AGCAGC--AGGTAAAGCT-GCTG-----CTCT 231
Db 670 AAGTAGTAGAGTAGAGGAGGAGAGAGTGGCTGAGTGGAGAGAGAGCGGAT 729

QY 232 -ACGACATGAAC-----GGCTGTACTAC----- 255
Db 730 GATGACGAGGACGATGAGGATGAGTGTGATGAGTGAAGAGAGGCTGAGGAACCTACGAA 789

QY 256 ---CGCTCAAGGAG-----CTGG---TGCCACCTGCGC-CCAGAACCGCA-AG---GTG 299
Db 790 GAAGCCACAGAGAACCCAGCATTTGCCACACACACACACACACAGAGTCTGTG 849

QY 300 AGCAAGGTG---GAGATTCT-----CCAGCAC-----GTCAATCGAC---TAC 335
Db 850 GAAGAGGTGTTTCAGTTCTTACAAAGCAGCAGTACCCCTGATGCCGTGACAAATAT 909

QY 336 ATC-AGGACCTTCAGTTGGAGTGAAC-----TC---GGAATCCGAG-----TT 377
Db 910 CTCGAGACACCTGGGATGAGATGAATGCAATGCCCATTTCCAGAAAGCCAAAGAGAGCCT 969

QY 378 GGGACC-----CCCGG----- 389
Db 970 GAGGCCAGCAGCAGAGAGAGATGTCAGTCTATGAGAGATGGGAAGGAGGAGAACGT 1029

QY 390 -----GGCCGAGGCTG-----CCGG-----TCCGG----- 411
Db 1030 CAAGCAAGAACTTTCCTAAAGCTGATGAAGAGCAGTTATCCAGCATTTCCAGGAGAA 1089

QY 412 -----CT-----CCGC-----TCAGC-----ACCTCA-- 429
Db 1090 GTGGATCTTTGACAGAGAGCAGCAACGAGAGACAGCAGCTGTTGGAGACACATG 1149

QY 430 -----ACGGCAGATCAG-----CGCC-CTGAC-GGCCGAGG----- 459
Db 1150 GCAGAGTGAAGCCATGCTCATATGACCGCGCGCGCTGGCCCTGGAGACTACATCACC 1209

QY 460 ---CGGATGC-GTTCCTGGGACGATCGCATCTTGT-----CGCTGAA----- 501
Db 1210 GCTCTGAGGCTGTTCTCTCC-TCGGCCTCGTCACTGTTCATATGCTAAAGAGATG 1267

QY 502 ---GCGC-----CTCCCCCAGGAGC-----GG 521
Db 1268 TCCGCGCAGAACAGAGGACAGACACACCTTAAGCATTTTCGAGCATGTGGCATGG 1327

QY 522 CGGACCCC-----AGCC-----ATCCAGG-----GGCA-----AGAG-GAATT 554
Db 1328 TGGATCCCAAGAAAGCGCTCAGATCGGTCCAGGTTATGACACACCTCCGTGTGATTT 1387

QY 555 ACCTGC-----TCTGTGGTCTCC-----CCACGCGCT-C-----GCCG----- 590
Db 1388 ATGAGCGCATGATCACTCTCTCTCCCTGCTCTACACGTCCTGAGTGGCGGAGAGA 1447

QY 591 -----GATCTG-----AGGAGAACAA-----GA 609
Db 1448 TTCAGGATGAAGTTGATGAGCTGCTTCAGAAAGAGCAAACTATTTCAGATGAGCTCTGG 1507

QY 610 CC-----GATCG-----CGCCAC-----TGGCG----- 629
Db 1508 CCAACATGATTAAGTGAACCAAGGATCAGTTACGAAACGATGCTCTCATGCCATCTTTGA 1567

QY 630 CTTTAAGTGCATCAGGCTGGGCT-----GAGGCTGAG-----CA---CTGG----- 670
Db 1568 CCGAAACGAAACCAACCGTGGAGCTCTCCCGTGAATGAGAGTTTCAGCTGACGATC 1627

QY 671 -----CGAGAG-----AGGGCGCT-CTCTCTGACACCTACT-AGTCACC-----AG 712
Db 1628 TCCAGCGGTGCACTCTTTGGGCTGACTCTGTGCCAGCAACACAGAAACCAAGTTG 1687

QY 713 AGACTTTAGGG-----GTGGGAT-TCCACTCG-----TGTGT 744
Db 1688 AGCTGTGTATGCCCGCCCTCTCGCAGCAGGAGTACCCACTCGACCGAGTTCTGGT 1747

QY 745 TCTATTTTTTGAAGACAGACATTTTAAATAATGTCACGTTTGGTGC-----TTCTCAG 800
Db 1748 TGACAAATATCAAGACGAGGAGATCTCTGAAGTG---AAGATGGTGCAGAAATCCGAC 1804

QY 801 ATTTCTGAGGAAAT--TGCTTTGATGATATATATATATATATATATATATATATATAT 858
Db 1805 ATGACTCAGATATGAAGTTTCATCAAAATTT--GGTGTCTTTG-CAGAAATGTTG 1861

QY 859 TTT--TACAAATAGTTC-----TGCGGCTGTTTTTT-----TGT 891
Db 1862 GTTCAACAAAGGTGCAATCATTTGGAATCATGTTGGGCGGTGTTGTTCATACGACAGTGA 1921

QY 892 TATTAACAAATAATTTAGATGGTGAATAAAAAA 925
Db 1922 TCTTATCACCTTGGT--GATGCTGAAGAGAAA 1953

RESULT 10
US-09-551-853D-13
; Sequence 13, Application US/09551853D
; Patent No. 6500667
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 29915/6280L
; CURRENT APPLICATION NUMBER: US/09/551,853D
; CURRENT FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23

```

; Patent No. 6420534
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; FILE REFERENCE: 29915/62801
; CURRENT APPLICATION NUMBER: US/09/548,372D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 2088
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-548-372D-13

Query Match      41.4%; Score 2385; DB 4; Length 2088;
Best Local Similarity 37.6%; Pred. No. 2.7e-100;
Matches 600; Conservative

QY 1 GGGGCCCATCTCT-GTTTCAG-CCAGT-----CGCCAGA-ATCATGAAG-TCG 45
DB 373 GATGCCCC-TTCTGTTCTCTCAAGTGCATAATTTCTTACACAGGAGGATGGATTG 431
QY 46 CCAAGTGCAGACCGC-----CACCGCGCGCGGCGCCAGCTGCGCGCTGAAGCCG- 99
DB 432 CGAACTCATCTTCACTGGCAGCAGCGCTCGCCAAAGAGACATGCACTGAGAGAGTACCAA 491
QY 100 ---GCAAGACAGG-----AGCGGTGCGGG-----CGAGGTG-----128
DB 492 CTTGCATGACTACGCGATGTTGCTGCTCGCGAATTGACAAGTTCCGAGGGGTAGATT 551
QY 129 ---GTG-----CGGTGCTG-----TCTG-----AGCAGAGCGTGG 156
DB 552 TGTGTTGTTGCCACTGCTGAAGAAAGTGAATGTTGATTCTGCTGATGCGGAG-GAGG 610
QY 157 CCATCTCGCGCTG-CCGCGCGCGCGGCGCGG-CTGCGTCCCTCTGCGAG- 207
DB 611 ATGACTCG-GATGCTCTGTTGGTGGCGGAGGAGACACAGACTATGCAATGGAGTGAAGAC 569
QY 208 -----AGCAGC--AGGTAAAGCT-GCTG-----CTCT 231
DB 670 AAAGTAGTAGAGTAGCAGAGGAGGAGAGAGTGGCTGAGGTGGAAGAGAGAGCCGAT 729
QY 232 -ACGACATGAAC-----GGCTGTTACTAC-----255
DB 730 GATGACGAGGACGATGAGGATGGTGTATGAGGTAGAGGAGAGGCTGAGGACCCCTACGAA 789
QY 256 ---GCCCTCAAGGAG-----CTGG---TGCCACCCCTGCC-CCAGAACCGCA-AG---GTG 299
DB 790 GAAGCCACAGAGAGAACCAACAGCATTGCCACCAACCAACCAACCAACAGAGTCTGTG 849
QY 300 AGCAAGGTG-----GAGATTCT-----CGAGAC-----GTCATCGAC---TAC 335
DB 850 GAAGAGGTGTTTCAGATTCTCTTACAAACAGCAGCCAGTACCCCTGATGCCGTGCAAGAT 909
QY 336 ATC-AGGACCTTCAGTTGGAGCTGAAC-----TC--GGAATCCGAAG-----TT 377
DB 910 CTCAGACACTGGGGATGAGATGACATGCCCATTTCCAGAAAGCCAAAGAGAGGCTT 969
QY 378 GGGACC-----CCGGG-----389
DB 970 GAGGCCAAGCCAGAGAGAGATGTCAGGTCATGAGAGATGGGAAGGCGAAGCT 1029
QY 390 -----GGCGAGGCGT-----CCGG-----TCCGGG-----411

```

```

DB 1030 CACGCAAGAACTTGCCTAAGCTGATAGAAGGCGATTATCCAGCATTTCCAGGAGAA 1089
QY 412 -----CT-----CCGC-----TCAGC-----ACCTCA--429
DB 1090 GTGGAATCTTTGGAACAGAGAGAGCCACAGAGAGAGAGAGTGTGGAGACACATG 1149
QY 430 -----ACGCGAGATCAG---CGCC-CTGAC-GGCGGAGG-----459
DB 1150 GCCAGAGTGAAGCCATGCTCAATGACCGCGCGCTGGCCCTGGAGAACTACATCACC 1209
QY 460 ---CGGCAATG-FTTCTCGGAGCGATCGCATCTTGT-----CGCTGAA--501
DB 1210 GCTCTGAGGCTGTTCTCTCC-TCGGCTCGTCAATGTTCAATGTTCTAAAGAGATG 1267
QY 502 ---CGGC-----CTCCCCAGGAGC-----GG 521
DB 1268 TCCGCGGAGAACAGAGAGAGAGACACACCTTAAAGCATTTCCAGCATGTGGCATGG 1327
QY 522 CGGACCCC-----AGCC-----ATCCAGG-----GGCA-----AGAG-GAATT 554
DB 1328 TGGATCCCAAGAAAGCGCTCAGATCCGCTCCAGGTTATGACACACCTCCGTGTGATTT 1387
QY 555 ACCTGC-----TCTCTGGGTCTCC-CCACGCGCT-C-----GCCG-----590
DB 1388 ATGAGCGCATGATCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1447
QY 591 -----GATCTG-----AGGAGAACAA-----GA 609
DB 1448 TTCAGGATGAAGTTGATGAGCTGCTTCAGAAAGAGCAAACTATTTCAGATGAGCTCTGG 1507
QY 610 CC-----GATCGG-----CGGCCAC--TGCGC-----629
DB 1508 CCAACATGATTAGTGAACCAAGAGTCACTTACGGAACAGATGCTCTCATGCCATCTTGA 1567
QY 630 CTTTAAGTGCATCAGCGCTGGGCT-----GAGGCTGAGG-----CA-CTGG-----670
DB 1568 CCGAAACGAAACACCGTGGAGCTCTTCCGCTGAATGAGAGTTTCAGCTGAGCAGATC 1627
QY 671 -----CGAGGAG-----AGGCGCT--CCTCTCTGCACACTACT-AGTCACC-----AG 712
DB 1628 TCCAGCGGTGCACTCTTTGGGCTGACTCTGTGCCAGCCCAACAGAAACCGAAGTTG 1687
QY 713 AGACTTTTAGGGG-----GTGGGAT-TCCACTCG-----TGTGT 744
DB 1688 AGCTGTTTGTATGCTCCGCTCTGCGAGCGAGGACTGACCACTCGACAGGTTCTGGT 1747
QY 745 TTCTATTTTGAAGAGCAGACATTTAAATAATGGTCACTGTTGGTGC-----TTCTCAG 800
DB 1748 TGACAAATATCAAGCAGGAGAGATCTCTGAAGTG---AAGATGATGCAAGATTCGAC 1804
QY 801 ATTTCTGAGGAAAT--TGCTTTGATTGATATTAATGATCACTACCGAGTGAATATTG 858
DB 1805 ATGACTCAGGATATGAAGTTCAATCAAAATTT--GGTGTCTTTG-CAGAAGATGTGG 1861
QY 859 TTT--TACAATAGTTC-----TGCGGCTGTTTTT-----TGT 891
DB 1862 GTTCAAAACAAAGGTGCAATCATTTGGGCGGTGTTGTCTATGCGACAGTGA 1921
QY 892 TATTAACAAATAATTAGATGCTGAAATAAA 925
DB 1922 TCTTCATCCTTGGT--GATGCTGAAGAGAA 1953

```

```

RESULT 9
US-09-548-367D-13
; Sequence 13, Application US/09548367D
; Patent No. 6440698
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; FILE REFERENCE: 29915/62801

```

Db 2303 CTACTGTACAAATCCAAAGCGTCTGCCGTATATTAGTGAACAGAGGGTGACCGTTGGCAT 2362
Qy 792 GCTTC-----TCA--GATTCTGAGGAAATGCTTTGTTATTTGTTATTTACAAATGATCAC 843
Db 2363 GCTCCAGACCTTCATTAAATCTCAATCAACTTCCCATCATTTCCATTTCGAATATCTC 2422
Qy 844 -----CGACTGAGAAATTTGTTTACAAATAGTCTGTGGGGCTGTTTTT 888
Db 2423 TTTTCTTGCACAGAGCGCTTTTTCGTTTTTTTATTGATTTATTTTACGATTTTT 2482
Qy 889 TGTATTAAACAAA-----TAATTAGATGTTGAAAAAAAAA 926
Db 2483 AGATAATGCACAGATGCTTCATTGCTCAATAAATTTATTTTAAATAAAAAAAA 2542

RESULT 7

US-09-436-063C-3
; Sequence 3, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095ccorrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-3

Query Match 41.6%; Score 2397; DB 4; Length 2544;
Best Local Similarity 38.7%; Pred. No. 8.9e-101;
Matches 580; Conservative 0; Mismatches 314; Indels 606; Gaps 83;
Qy 2 GGGCC-----CATTCTGTTTC-----AGCC-----AGT- 24
Db 1074 GGTCTGGAGTGTCTTCTTAATTTTTCGTAATGTTTCGCTCGTCTCGATGTA 1133
Qy 25 CGCCA-AGAAT---CATGAA-AGTCG-----CCAGTGGCAGCA-----C 58
Db 1134 CGCCATAGTATCTACATGAATAACGACTGGTCTCCGACGGGAAACGAAGAAAGC 1193
Qy 59 CGCC--ACCGCCCGCGGGCC-----CCAG-----CTGCGCG---CTGAAGCGC----- 99
Db 1194 CGCCGAACACACAGCAGCGAAACGAGATGCCAATGTTTCAACGCGAGCGCGCGCCAA 1253
Qy 100 -----GCAAGACA-----GCCAG 112
Db 1254 TAATAATGCTGATTTGATTTGCCGACACAAATTCCTCTATGATCCATTGATGAGAT 1313
Qy 113 C-----GGTG-----CGGCGA-----GGTGGTGGCGCTGC-----TGTCTGAG- 146
Db 1314 CCCAGAAAATTTGATTTGCGGCGAGATTCCAATGATGCAACATCCACGTCTTTGTACAGA 1373
Qy 147 CAGAGC-----GTGGCCATC-----TCGCGCTGCCG-----GGCG--GCCGGCGCGC 187
Db 1374 CGCGGCACATACGCTATGCGCGGCTCCATTGCGCGCGCGCGGCGGCGGCGGCGGCGGCG 1433
Qy 188 CTTGCTGC-----CTGC-----TGAACGAGC--AGCAGGTA-----AACGT 223
Db 1434 -CTGCCAAGCATGAGCGCTGCAAAAATCGATAAGCTTAGCGGATACGTTTCCCATTTGT 1492
Qy 224 GCTGCTCTA---CGACATGACGGCTG-----TTACTCAGC 257
Db 1493 CTTTCTCTATCTCAATAGTCTACTGTTGTATATGAAATATCTAAGCTTAAACTCGT 1552

Qy 258 C--TCAAGGAGCTGTGSCCACCCT---GCCCGAGAACCCGAAGGTGAGCAAGGTGAGA 312
Db 1553 CGGACAAAGATCCAGGAGAACGACAAAGTGGCAGCAGATCCAC-----TGATCGGTA 1602
Qy 313 TTCTCCAGCAGCTCATCGACTACATCAGG--GACTTCAGTTGGAGCTGAACTCGAATCC 371
Db 1503 TTCAGCGC--CGAAATCGAGTACAAATGTTGTAGCTC--GAAGAGCGCGAAATTTGTCACA 1660
Qy 372 GAAGTTGGAGCCCGCGGGCGCGAGGGCTGCCG-----GTC-- 407
Db 1661 GCGGTCAAGCGCGACGCGAATCGAAATCTCGAGTTATAAATTCATAAAATCTGCCAA 1720
Qy 408 ---CGGCTCGCTCAGCAC-----CC-----TCAACGG----- 433
Db 1721 AACGGACACTTGCACGACTTCATCGGGGACCTACTCTCTCTACGGGTAGTTTCATA 1780
Qy 434 -----CGAGATCAGCG-----CCCTGACGGCC----- 455
Db 1781 TTTGATCGGA--CAGCGGCTTCTACTTTCTTCAAAATATTTTCCCTGCCAGCGCTCGTCT 1839
Qy 456 -----GA---GGCGGCAT--GGTTCC----- 472
Db 1840 AGTTTATATGATGATCTCTGATCAATTCGATCAATCGGCTCGCGCTTTCGCGAACCCTAAT 1899
Qy 473 ---TGCG--GACGAT-----CGCATCTTGTCTCGCTGAAGCG-----CCTC 508
Db 1900 CGGTACGATGACGCTGCTCACTGAGACTCACTTATGAC--CGGAACCAATCGACGTCTTC 1958
Qy 509 CCCAGGAGCGGGG-----GAC-----CC-----CAGCATC--CAGS--- 540
Db 1959 CACCAGTTGCTTATGTAAGAGCGCTTGATGATTTCTCTCGGTTCTTCTGATCTCTCTGTTA 2018
Qy 541 ---GGGC--AAGAGGAATTCGTCTCTGTG--GCTCTCCCCC-----AACGGCGCTC 586
Db 2019 TACTGGGTTGATCGAGTAGC---CCTGTGTTCCCTACTCAAAAAGAGAGACGAGATC 2075
Qy 587 GCCGATCTGAGGAGAACAGACCGATCGCGGCC--ACTGCGCC-----CT 632
Db 2076 GTCCGA--GAAGAGAGAAAGACGAGCATAAACCTGTCTCCCGCACACCTGATATCT 2133
Qy 633 TAATGCTATCCAGCTGGGCTGAGGCTGAGGCTGAGGAGGAGGAGGCGC--TCTCT 690
Db 2134 TCAGCAGCTC--GGCT--TGCCGA-----ATGCATGCAACGC---GGCTCAACCTCG 2182
Qy 691 CTCACACC-----TACT-----AGTCAC-----CAGAGACTT- 718
Db 2183 ATCATCGCGCTCATCAAGCAGTCCGAATCGATTCGTGTGATCAGTCACAGTCACATTCATC 2242
Qy 719 -TAGGGGT---GGATTCACATCTGT-----GTGT-----TCTATTTTTC- 756
Db 2243 GTACGCGTGGCGGTTTCTCTTGTGTTTCACTTGTCAACACTCTCTTCTGGTGATT 2302
Qy 757 -----AAAAAGCAGACATT-----TTAAAAAATGGTCACG-----TTTGGT 791
Db 2303 CTACTGTACAAATCCAGCGCTCTGCCGTATATTAGTGAACAGAGGGTGACCGTTGCGAT 2362
Qy 792 GCTTC-----TCA--GATTCTGAGGAAATTCGTTTGTATTTGTTATTTACATGATCAC 843
Db 2363 GCTCCAGACCTTCATTAAATCTCAATCCACTTCTCATCTTTCATTTTCCATTTGGAATATCTC 2422
Qy 844 -----CGACTGAGAAATTTGTTTACAAATAGTTCTGTGGGGCTGTTTTT 888
Db 2423 TTTTCTTGCACAGAGCGCTTTTTCGTTTTTTTATTGATTTATTTTACGATTTTT 2482
Qy 889 TGTATTAAACAAA-----TAATTAGATGTTGAAAAAAAAA 926
Db 2483 AGATAATGCACAGATGCTCTCATTCCTCAATAAATTTATTTTAAATAAAAAAAA 2542

RESULT 8

US-09-548-372D-13
; Sequence 13, Application US/09548372D

QY 623 -ACTGGCC-----CTTAAGTGCATCAGCTGGGCTGAGGCTGAGGCACT 668
DB 2074 TGCTCGCGGACACCTGATATCTTTCAGAGTCC-GCCT--TGCCGA-----ATGACA 2125
QY 669 GCGAGGAGAGGCGCG--TCCTCTCTGCACACC-----TACT-- 703
DB 2126 TGCAAGCC--GCTCCAACTCGATCATCGCGTCATCAAGCAGTCCGAATCGATTCTGT 2182
QY 704 ---AGTCAC-----CAGAGACTT--TAGGGGT--GGGATTCCACTCGT-----GTG 743
DB 2183 GTCAGTCAAGTCAATGATGATCGTCAAGCGTTCCTCTGTGTTTCACTTG 2242
QY 744 TT-----TCATTTTGTG-----AAAGCAGACATTT-----TAAATAAT 778
DB 2243 TTCAACACTCTCTTCTGGTGTATCTACTGTACAAATCCAAAGCGTTCGCGTATATAGT 2302
QY 779 GGTACAG-----TTTGGTCTTC-----TCA--GATTCTGAGAAATTCCTT 819
DB 2303 GAACAGAGGGTCAAGCTTGCATGCTCCAGACCTTCATTAATCTCAATCCACTTCCTC 2362
QY 820 TGATTTGATATATCAATGATCAC-----CGACTGAGAAATTTGTTTTTAC 864
DB 2363 ATCATTTTCCATTTCCGATATCTCTTTTCTTTCACAGAACCTTTTTCGTTTTTTTTT 2422
QY 865 AATAGTTCTGTGGGCTGTTTTTTTGTATTAAACAAA----- 902
DB 2423 ATGATTTATTTTACGGATTTTATAGATAATGCCAGATGCTCATTTGCTCAATAAAT 2482
QY 903 TAATTTAGATGGTGAATAAAAAA 926
DB 2483 TATTTTAAATTAATAAAAAA 2506

RESULT 6
US-09-627-650B-3
; Sequence 3, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Erik
; APPLICANT: Bamber, Bruce
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.000903
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-3

Query Match 41.6%; Score 2397; DB 4; Length 2544;
Best Local Similarity 38.7%; Pred. No. 8.9e-101;
Matches 580; Conservative 0; Mismatches 314; Indels 606; Gaps 83;
QY 2 GGCC-----CATTCTGTTTC-----AGCC-----AGT- 24
DB 1074 GGGTCGGAGTGTCTTATTTTGTGTTTGTATGTTATCGCTCGCTGTCGAGTA 1133
QY 25 CGCCA-AGAAAT---CATGAA-AGTCG-----CCAGTGGCAGCA-----C 58
DB 1134 CGCCATAGTATCTACATGAATAAACGACTGCTCCGCGGAAACGAGAAAGC 1193
QY 59 CGCC--ACCGCGCGCGGGCC-----CCAG-----CTGCGCG--CTGAGGCGG--- 99
DB 1194 CGCCGAACACAGCAGCAAGATGCCATGTTTCAACCGAGCGCGGAGGCGCCAA 1253

QY 100 -----GCAAGACA----- 112
DB 1254 TAATAATGCTGACTTGTACTTTTGGGACACAAATTCCTCTATGAAATCCATTGATGAGAT 1313
QY 1113 C-----GGTG-----CGGCGCA-----GGTGGTGGCTGTC-----TGCTGAG- 146
DB 1314 CCCGAAAAATTTGATTGCGCGGACGATTCGAATGATGCAACATCCACGCTTGTGTACAGA 1373
QY 147 CAGAGC-----GTGGCCATC-----TCGGGCTGCGG-----GGGC--GCGGGGCGG 187
DB 1374 GCGGCGCATACGCTATGCGGGCTCCATTCGCGCGCGCGGAAAGAGCTTCCAGACATG 1433
QY 188 CTTGCTGTC-----CTGC-----TGACGAGC--AGCAGGTA-----AACGT 223
DB 1434 -CTGCCAACGATGGACGCGCTGCAAAATCGATAAGCTTAGCCGATACGGTTTCCCATTTG 1492
QY 224 GCTGCTCTA-----CGACATGAACGGCTG-----TTACTCAGC 257
DB 1493 CTTTCTCTATCTTCAATATAGTCTACTGTTGTATGAAATATCTAAGCTTAACTCT 1552
QY 258 C--TCAGAGAGCTGGTCCCACTCT--GCCCGAGAACGCAAGGTGAGCAAGGTGAGGA 312
DB 1553 CGGCAAGATCCAGGAGAGCAAGTGGCAGCAGATCCAC-----TGATGCGTA 1602
QY 313 TTCTCCAGCAGCTCATCGACTACATCAGG--GACCTTCAGTTGGAGCTGAACCTCGGAATCC 371
DB 1603 TTGACGCGC--CGAATCGAGTACAAATGGTGTACGTC--GAAAGAGCGGATTTGTCGACA 1660
QY 372 GAAGTTGGAGCCCGCGGGCGGAGGGCTGCGG-----GTC-- 407
DB 1661 CGGTCAAGGCGGACGCAACATCGAACTGTGAGTTATAAATCTACTAAAAATCTCCAA 1720
QY 408 ---CGGGCTCGCTCAGCAC-----CC-----TCAACGG----- 433
DB 1721 AAACGACACTTGCAGCAGCTTTCATCGGGGACCTACTCTGCTCTACGGGTAGTTTCA 1780
QY 434 -----CGAGATCAGCG-----CCCTGACGCGC----- 455
DB 1781 TTGATCGCGA--CAGCGGCTTCTACTTTTCAAAATATTTTCCCTGCGAGCTCTGCT 1839
QY 456 -----GA-----GGCGGCAAT--CGCTTCC----- 472
DB 1840 AGTTTATATGATCTCATTTCTGGATCAATCGTACTCGGCGCTTCGGAACCTTAAT 1899
QY 473 ---TGCG--GACGAT-----CGCATCTTGTGCTGCTGAAGG-----CCCTC 508
DB 1900 CGTACGATGACGCTGCTCTGAGCTCATCTTATGAC--CGGAACCAATCGACGCTCTC 1958
QY 509 CCCAGGAGCGCGC-----GAC-----CC-----CAGCCATC--CAGG--- 540
DB 1959 CACCACTGCTATGTAAGGCGGTTGATGTTTCTCGGTTTCTGCTATCTTCTGTTA 2018
QY 541 ---GGGC--AAGAGGAATTAGTCTCTGTG--GGTCTCCCGC-----AAGCGCGCTC 586
DB 2019 TACTGGGCTTATCGAGTACG-----CCTGTTTCTCTACTCAAAAAGAGACGAGATC 2075
QY 587 GCCGATCTGAGGAGAACCAAGACCGATCGCGGCGC--ACTGCGC-----CT 632
DB 2076 GTCGGA--GAAGAGAGAAGACGAGCATAAACCTGTCTCCGCCACACCTGATATCT 2133
QY 633 TAATGCTATCCAGCTCGGGCTGAGGCTGAGGCTGAGGAGGAGGAGGCGC--TCCTCT 690
DB 2134 TCAGAGCTCC-GCCT--TGCCGA-----ATGACATGCAACG-----GGCTCAACCTG 2182
QY 691 CTGACACCC-----TACT-----AGTCAC-----CAGAGACTT- 718
DB 2183 ATCATCGCGCTCATCAAGCAGTCCGAATTCGATTTCTGTGTCAGTCCACATTCATGATC 2242
QY 719 -TAGGGGT--GGGATTCCACTCGT-----GTGTT-----TCTATTTTTTG----- 756
DB 2243 GTCAGCGTGGCGGTTCTCTTGTGTTTTCATCTTGTTCACACTCTCTTCTGGCTGATT 2302
QY 757 -----AAAGCAGACATTT-----TAAAAAATGGTCACG-----TTTGGT 791

1780 ATTTTCCCTGCCAGCCTCGTGTAGTTTATCATGATCTCATTTCTGATCAATCGTGA 1839
458 ---GGGGCAT-GGTTTC---TGCG--GAGCAT---CGCATCTTGT 492
1840 CTGGGCGCTTCGGAACTTAATCGGTACGATGACGGTGTCTCACTGAGACTCACTTAT 1899
493 GTGCTGAAGCG-----CTTCCCGCCAGGACCGGCG-----GAC-----CC 528
1900 GAC-CGGAACCAATCGAGCTCTTCCACAGTTCCTATGTATTAAGCGGTGTATTC 1958
529 -----CAGCCATC-CAGG-----GGGC-AAGAGGAATTACGTCTGTG-GGTCT 571
1959 TCGGTTCTGCTATCTTCTGTGTATCTACTGGCTGTGATCGAGTACG---CCTGTGTGCT 2015
572 CCCCC-----AACCGCCCTCGCGGATCTGAGGAGAAACAAGCCGATCGCGGCG 622
2016 ACTCAAAAAGAAAGACAGGATCGTCGGA--GAAGAGAGAGAGAGCGGATTAACC 2073
623 -ACTGCGCC-----CTTAATGCTATCCAGCTCGGCTGGGCTGAGGCTGAGGCACT 668
2074 TGCTCCCGCCAGACCTGATATCTTACGAGCTCC-GCCT--TGCCGA-----ATGCACA 2125
669 GCGGAGAGAGGGCGC--TCTCTCTGCACACC-----TACT-- 703
2126 TGCAACGC---GGCTCCAACTCGATCATCGCGCTCATCAAGCAGTCAATCTGT 2182
704 ---AGTAC-----CAGAGACTT--TAGGGGT---GGGATTCGACTCGT-----GTG 743
2183 GTCAGTCACAGTCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2242
744 TT-----TCTATTTTTC-----AAAAGCAGACATTT-----TTAAAAAAT 778
2243 TTCAACACTCTTCTGCTGATTTCTACTGTACAAATCCAAGCTGTGCGGTATATTAGT 2302
779 GGTCAAG-----TTTGTGCTTC-----TCA--GATTTCTGAGGAATTTGCTT 819
2303 GAACACAGGGGTGACCGTTGCGATGCTCCAGACCTTCATTAATCTCAATCAACTTCCTC 2362
820 TGTATTGTATATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864
2363 ATCAATTTCAATTTGATATCTTTCTTCTGACAGAGCTTTTTCGTTTCTTTT 2422
865 AATAGTTCTGTGGGCTGTTTTTTTGTATTAACAAA----- 902
2423 ATTGATTTATTTTACGGATTTTATAGATAATGACAGATGCTTCATTTGCTCAATAAAT 2482
903 TAATTTAGTGTGCAAAAAAAA 926
2483 TATTTTAAATTAATAAAAAAAA 2506

RESULT 5
US-09-436-063C-7
; Sequence 7, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PKT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-7

Query Match 41.8%; Score 2412; DB 4; Length 2508;
Best Local Similarity 39.5%; Pred. No. 1.9e-101;
Matches 578; Conservative 0; Mismatches 316; Indels 570; Gaps 81;
QY 2 GGGCC-----CATTCGTGTTTC-----AGC-----AGT- 24
Db 1074 GGGCTGATGTGTTTCTTAATTTTGTTCGTAATGATATCGCTCGTGTCTCGAGTA 1133
QY 25 CGCCA-AGAAT-----CATGAA-AGTCG-----CCAGTGGCAGCA-----C 58
Db 1134 CGCCATAGTATCTTACATGAATAAACGACTGCTCCGACGGGAAAAACAAGAAAGC 1193
QY 59 CGCC--ACCGCGCGCGGGGCG-----CCAG-----CTGCGCG--CTGAGGCGCGCAA 103
Db 1194 CGCCGAACAACAGCAGCGAAACGAGATGCCAATGTTCAACCGAGCGCCGAGGCGCGCAA 1253
QY 104 GA-----CAGCGAGCGGTG-----CGGCGCA-----GTTGTT 130
Db 1254 TAATAATAATCAATTGATGGAGATCCCAAGAAATTTGATTGCCGAGCAATTCATATGAT 1313
QY 131 GCGGTGTC-----TGCTGAG-CAGAGC-----GTGGCCATC-----TCGCGCT 168
Db 1314 GCAACATCCAGCTTTGTACAGACGGCGCACATACGCTATGGCGCGCTCCATTCGCGCG 1373
QY 169 CCGC---GGGC-GCGGGGCGCGCTGCTGCTG-----CCTGC-----TGACCGAGC 210
Db 1374 GCGGAAAGAGCTTCCAGACATG-CTGCCAACGATGAGCGCTGCAAAATCGATAAGC 1432
QY 211 -AGCAGGTA-----AACGTGCTGCTCTA-----CGACATGAAACGGCTG----- 247
Db 1433 TTAGCCGATACGTTTCCCATTTGCTTTCTCTATCTTCAATATATGCTACTGTTGTATA 1492
QY 248 -----TTACTCAGCC--TCAAGGAGCTGTGCCCAACCT--GCCCCAGA 288
Db 1493 TGAATATCTAAGCTTAACTCTGCGACAAAGATCCAGAGAACGACAGTGGCAGCAGA 1552
QY 289 ACCGCAAGGTGAGCAAGTGGAGATTCTCAGACGTCATCGATCATCATCAGG-GACCTT 347
Db 1553 TCCAC-----TGATGCTATTTCAGCGC-CGAAATCGAGTACAAATGGTGTAGCT 1601
QY 348 CAGTTGGAGCTGAACTCGAATCCGAATCGAAGTTGGGACCCCGGGGCGCGAGGCTCGCG-- 404
Db 1602 C-GAAGGAGCGCAATTTGTCAGACGCTCAAGCGCCGACCGCAACATCGAACTGTGAGT 1660
QY 405 -----GTC-----CGGCTCCGCTCAGCAC-----CC--- 426
Db 1661 TATAAATTCATAAAATCTGCAAAAACGGACACTTGCAGCACTTCATCGGGAGCCTAC 1720
QY 427 -----TCAACCG-----CGAGTCAGCG----- 444
Db 1721 TCTGCTACGGGTAGTTTCATATTTGATCGCGA-CAGCGGCTTCTACTTTCTTCAAT 1779
QY 445 -----CCCTGACGCGC-----GA 457
Db 1780 ATTTTTCCTGCGAGCCTCGTGTAGTTTATCATGATCTCATTTCTGATCAATCGTGA 1839
QY 458 ---GGCGCAT-GGTTTC-----TGCG-GACGAT-----CGCATCTTGT 492
Db 1840 CTCGGCGCTTCGCGAACCCCTAATCGGTACGATGACCGTGTCTACTGAGACTCATCTTAT 1899
QY 493 GTCGCTGAAGCG-----CCTCCCGCAGGACCGCG-----GAC-----CC 528
Db 1900 GAC-CGGAACCAATCGAGCTTCTCCACAGTTGCTATGTAAGCGCTGTGATGATTC 1958
QY 529 -----CAGCCATC-CAGG-----GGGC-AAGAGGAATAGTGTCTGTG-GGTCT 571
Db 1959 TCGGTTTCTGTATCTTCTGTGTTTACTGCGGTTGATCGAGTACG--CCTGTGTGCT 2015
QY 572 CCCCC-----AACCGCTCGCGGATCTGAGGAGAACAGACCGGATCGCGGCG 622
Db 2016 ACTCAAAAAGAAAGACAGGATCGTCGGA--GAAGAGAGAGAGAGCGGAGCATAAACC 2073

229 TCTACGACATG---AACGCTGTACTC-----ACG-----CCTCAAGGA 265
Db TCTGTCCGTGGAGAGCGTCAGAGCCCTCGTGTGACGTGAGAGGGTCTCTATGA 476
266 GC---TGCTGCCACCTC---GCCCAAGAACCGCAAGGTGAG---CAAGGTGAGAT-TC 315
Db CCCTTCTCTGCCCTCTGAGACTCAGACCG---GGAGAGAAGCCTCGTGGAACTCTG 534
316 TCAGACAC---GTATC---GACTACATC----- 338
Db TACAACACCGGTGCGCATGTCCTTCTGCTGCGCCCGCGCTGTGTTAATGTGCT 594
339 --AGGACCTTCAGTTGGAGC-----TGAACT--CGGAATCCGAAGTTGGAGCC 384
Db GGGGGACCTCTCTTTATAGCCACCGACTCAGTGAACCTGCGGTGC--TATTGGAGCA 652
385 C---CGG---GGC-----CGAGGTGCGGTCCGG---CTCCGCT-----CAG 421
Db CGAGCGAGGTGCTGTGAACACAGATCAACC-ATCAGGGTTTCTCTGCTGAGGTGCA 711
422 CAC--CCTCAACGCG-----GAGA---TCAGGCGCTCGACGCGCGAGGC--- 460
Db CTCATCCACTTCAACCAAGAACTCTATGGAACTCTAGTGCC--GCCACGAGGGGCCA 769
461 --GGCATG--CGTTC-CTGC-----GGACGATCGC---ATCTTGTGCTGCT 498
Db ATGGCCTGGCCTCTCTCAGCCTCTTTGTCAANGTGTGCTAGCTCAAAACCCGTTCCCTCA 829
499 GAAGCGCTCC---CCACGGGACCGCGGACCC-CA-----GCCATC 536
Db GAG--CGCTCTTAAACGTCGACCATC--ACCGCATCTCTTAAGAATGATGCGCTAC 885
537 -----CAGG-----GGCAAGAGAAATAC-GTGCTCTGTG----- 566
Db TTTCTTCAAGACTGAGCC--TGAGCTCTGTGTCGCCCGAGTCTTTGGCTTCATCACT 943
567 ---GGTCTCCCAAGCGCTCGCGGATCTGAGGGAGAACAGACC--GATCGGG 620
Db ATCAGGCTCTCTCAGCACCCAC-CTGTCTC---GGAGCTGTACTGTGATCTCAT 998
621 CCACTGCGCCTTAA-----CTGCATCCAG-----CCTGGG---CTGAGGCTGA- 662
Db TGACAGGCGCTCAATATACCTCCCTCCAGCTGCATTCCTCGGACTTCTGAGCGAGAA 1058
563 -----GGCACTGGGAGAGAGGC-GCT-----CC--TCTCTGCACACTACTA 704
Db TCCTCGGGTCCAGTCTTCCAGAGCTCAGCCGTAAGCGCGCGCCCTGAGCCCTTCG 1118
705 GTCACAGAGACTTTAGGG--GTGGGATT-----CCACTCTGTG----- 743
Db G-CA-CAGGCGCTTGAGGGGCAACAGGACCCCGGCAACCGAGAGGGTGGAGGGCC 1176
744 TTTCTATTTTGAAGAGAGACATTTTAAATAATGTCACGTTTGGTCTCTCAGATT 803
Db CAACTACCGCTG-CATGTGATGATGTCCTCCCATGCTCTC---TGAGACTCCCCA- 1228
804 TCTGAGAAATGCTTTGTATTTATATTAACAATGATCACCGACTGAGAATATTGTTTA 863
Db TC-GAGGAGTCCG- ----CTGCTTTTCCCAAGCTCCCC-ACCAGGGGAGGGAGTG 1280
864 C-----AATAGTTCTGT-----GGGCTGTTTTTTTGTATTAACAATAATTAGA 911
Db CACCCCAACAAAGCTATTAAAGGAGACAGATACTTCTCTGTAAAAA- 1334
912 TGGTGAATAAAAAA 926
Db -----AAAAA 1345

RESULT 4

US-09-627-650B-7

; Sequence 7, Application US/09627650B

; Patent No. 6406872

GENERAL INFORMATION:

; APPLICANT: Bauber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-7

Query Match 41.8%; Score 2412; DB 4; Length 2508;

Best Local Similarity 39.5%; Pred. No. 1.9e-101;

Matches 578; Conservative 0; Mismatches 316; Indels 570; Gaps 81;

QY 2 GGGCC-----CATTCGTTC-----AGCC-----AGT- 24
Db 1074 GGGTCTGATGTGTTCTTAATTTTGTTCGTAAGTATTGCTTCGTTCGTGTCGAGTA 1133
QY 25 CGCCA-AGAAT---CATGAA-AGTCG---CCAGTGGCAGCA-----C 58
Db 1134 GGCATAGTATCTACATGATAAAGACTGCTCCGACGGAGAAAACGAAAAAGC 1193
QY 59 GCGC--ACCGCGCGCGCGGCG---CCAG---CTGCGCG---CTGAGGCGCGCAA 103
Db 1194 GCGCAACAACAGCAGCAGCAACGAGATGCCAATGTTCAACGCGAGCCGAGGCGCGCAA 1253
QY 104 GA-----CAGCGAGCGGTG-----CGGCGCA-----GTTGGT 130
Db 1254 TAATAATAATCCATTGATGGAGATCCAGAAAATGTTGATGTCGCGAGTTCATCAATGAT 1313
QY 131 CGCTGTGC-----TGTCTGAG-CAGAGC---GTGGCATC-----TCGGCT 168
Db 1314 GCAACATCCACGCTTGTTCACAGCGCGCACATACGCTATGCGCGGCTCCATTCGCGCG 1373
QY 169 GCGC---GGGC-CCCGGGCGCGCTGCTGC---CCTGC---TGGACGAGC 210
Db 1374 GCCAAAAGGCTTCCAGACATG-CTGCCAACGATGGAGCGCTGCAGAAATCGATAAGC 1432
QY 211 --ACAGGTA-----AACGTGTGCTCTA---CGACATGAACGGCTG----- 247
Db 1433 TTAGCCGATACGCTTCCCATTTCTCTATCTTCTATCTTCAATATAGTCTACTGTTGTATA 1492
QY 248 -----TTACTCAGCC---TCAAGGAGCTGGTCCCAACCT---GCCCCAGA 288
Db 1493 TGAATATCTAAGCTTAACTGTCGACAAAGATCCAGGAGACGACAGTGGCAGCAGA 1552
QY 289 ACCCAAGGTGAGCAAGTGGAGATTCTCAGCAGCTGTCATCGACTACATCAGG-GACCTT 347
Db 1553 TCCAC-----TGATGCTATTTCGACGGC-CGAAATCGAGTACAAATGTTGTACGT 1601
QY 348 CAGTTGGAGCTGAACTCGGAATCGGAATGGGACCCCGGGGCGCGAGGCTGCGC--- 404
Db 1602 C-GAAGGAGCCGATTTGTCAGACGCGTCAAGCCGACCGCAACATCGAACTGTGAGT 1660
QY 405 -----GTC-----CGGCTCCGCTCAGCAC-----CC--- 426
Db 1661 TATAAATTAATAAATCTGCCAAAACGAGACATTCGACGACTTCATCGGGAGCCTAC 1720
QY 427 -----TCNACGG-----CGAGATCAGCG----- 444
Db 1721 TCTCGTCTACGGGTAGTTTCATATTGATCGCA-CAGCGGCTTCTACTTCTTCAAT 1779
QY 445 -----CCCTGACGGCC-----GA 457

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Yue, Henry
APPLICANT: Greenwald, Sara
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,767
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0423 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1532042
US-08-977-767-3

Query Match 42.2%; Score 2431.5; DB 2; Length 1345;
Best Local Similarity 42.2%; Pred. No. 1.6e-102;
Matches 564; Conservative 0; Mismatches 301; Indels 470; Gaps 89;

QY 156 -GCCATCT-CGCGC-----TGCC-----GGGGCGC-----CGGGGC-----GCGCCT 190
Db 417 CACCATCTACGCTGGATGCTGCTCTTTGGGGCCCTCGTCTCAAGGCGCGTGCACCT 476
QY 191 GC-----CTGCGCTGC--TGAGCAGCAGCAGGTAAACGT--GCTGC--TCTAC--GACA 237
Db 477 GCTCATCTACCTCACCATGTAGC-CAGCAGCTTTACGCTGCTGCTCTCCGTCGAC 535
QY 238 TGAAC--GGC--TGTTACTACGC-----CTCAAGAGCTGTGTGCCACACCTG--CCCA 286
Db 536 GGTACTCTGGCGTGGCGCACCGCTGCTGCGCGCCCTGGCGCACGCGCGTAAACGCC 595
QY 287 GAACCGCA-----AGGTAGCAAGGTGGAG-----ATTCTC--CAGCAGCTCATCGAC 332
Db 596 GCGCCGACGTGGGGCTGGTGTGGCTGCTGGCGCGCTCTTCTCGGC-GCCCTACTCTAGC 654
QY 333 TACATCAGGACCTTCAGTTG-----GAGCT-GAATCGGAAT--CCGAAGTTGGA----- 381
Db 655 TACTAC--GGCAGCTGCTACGGCGCTGGAGCTCTGCTGCGCGC--TGGGAGGAC 711
QY 382 -----CCCCGGG-----GGCGAGG-----GCTGCGG--TCGGGCTCGCTCAG-- 422
Db 712 GCGCGCGCGCGCCCTGGACGTGGCCACTTTCGCTGCGCGCTACTGCTGCGCGCTGGCT 771
QY 423 -----ACCTC--CAAC--GGCAGATCAGCGC--CCT--GACGGCCGAGG--CGGCA 464
Db 772 GTGGTAGCTGSCCTACGGCGCAGCTGCGCTTCTGTGGCGCGCGCTGGGTCCCGCG 831
QY 465 TGCGTCTCTCGGACGA-TGCGATCTTGTGTCTCTGAAAGCGC-----CTCCCC----- 511
Db 832 GCGCGCGCGCGCGCGAGGCT 891
QY 512 -CAGGACCGCGCGACCCC--AGCCATCCAGGGGGCAAGAGGAATACGTGCT--CTGT 565
Db 892 GCGGTGGCC--GCGCTTACCGCTCTGCTGGGTCC-----GCACACGGCGCTCATCTGT 946
QY 566 GGTGCT-----CCCCAACGC-----GCTC-- 586
Db 947 GCTTCTGTAGCGCGCTTCCCTTACGCGCGCCACCTACGCTCGCGCTGCGCTCAC 1006
QY 587 -----GCGGATCTGAGGAGAACAG-----ACCGATCG-----GCG--GCGCTGCGC 629
Db 1007 ACTGCTGCGCTACGCCAATCTCTGCTCAACCGCTGCTGTCTACGCGCTCGCTC--GCGC 1065
QY 630 CTTTAAAC-----TGATC--CAGCCTGGGCG--TGAGGTG--AGGCACCTGGGAGAG--A 678
Db 1066 CACTTCCGCGCGCTTCCGCGCGCTGTGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCT 1125
QY 679 GGGCGCTCTCTCTGACACCTACTACTACGAGACCTTTAGG-----GGGT----- 726
Db 1126 GCGCGCGCTTGTCTGCGCTCGC-----CCGCGCTCTCGGGGCCACCCGCTGCCCC 1179
QY 727 -GGGATTC--ACTCGTGTGTTCTATTTTGTAAAGCAGACA-----TTTTTAA 774
Db 1180 GGAGACCGCGCGCTAGCGGAGGCTGCTGGCTG--GTGGCGCGCGCGCGCGCGCGCGCG 1238
QY 775 AAATGGTACGTTTGGTCTTCTCAGATTTCTGAGGAAATGCTTTGTATTGTATTATAC 834
Db 1239 GGAGGACCGCTTCCACGCGGAGAGGCTGCGCGAGGACC-----GGAATAACCTGCTC 1291
QY 835 --AATGATCCCGACTGAGATAT--TGTTTACAATAGTTCT-----GTGGGG----- 879
Db 1292 GCGCTGACTCGCGCTGTGCTGCTGTCTCACTCCCGTCTCCGAGGCGGCGAGCGCCAC 1351
QY 880 -----CTGTTTTTTTGTATTAAACAAATATTAGATGTTGTAATAAAAAA 925
Db 1352 CGGGCAGGATGGGCAATGCCACGAGCTCTCTGAGGGCGGTGAGTGA 1402

RESULT 3
US-08-977-767-3
; Sequence 3, Application US/08977767
; Patent No. 5972684

Db 398 GATCCACTCTCTATCTGTCACCGGCTGATCTCAACAGGAGACAGCAAGATGAGTTCAAGG 457
QY 217 TAAAGCTG---CTGC-----TCTAC---GACATGACGCTGTACTACGCGC--- 258
Db 458 AGATCAAGAGCTGCTGGAATGCTCAACGTGACATGACATGACATGACCTACTCTCC 517
QY 259 ---TCAAGAGCTGGTCCAC-CCTGCCCGAGAACCG-CAAG-GTGAGC-AAGGTGGAG 311
Db 518 TCTTCAAGAGT--GTGACCACTCCAACAC--GACCGCTAGAGGGGGCTGAGATCGAG 573
QY 312 ATTCTCCAGC-ACGTC-----ATCGACTACATCAGG---GACCTTCAGTTGAGCTGA 360
Db 574 GAGTTCCTGCGGCGCTGCTGAAGCGCGGAGCTGAGGAGATCTTCCAT-----CAT 628
QY 361 ACTCG-----AATCG-AAGTTGGACCCCGGG---GCCAGGG---CTGCGGTTC--- 408
Db 629 ACTCGGCGAGACCGGTGAGTGCCTGAGTGCCTGAGCTGCTGAGTCTTCTGGAGGACGAG 688
QY 409 -----GGGCTCGCTCAG---CACCTCAACCGCGAGATCAG-----CGCCC 447
Db 689 GCGAGGAGGCGCCACACTGCGCGCGCCGCGAGCTCAATCAGACTATGAGCTCAAG 748
QY 448 TGACGCGGAGCGGCGCTGCTTCTGCGGAC-----GATCGCATCTTG-----TG 493
Db 749 AGACAGCC-AAGCAGCATGAG---CTGATGACATGATGCTTCAATGATGATCTGTTG 804
QY 494 TCGCTGAAG-----CGCCTCCG-----CCAG-----GGACCGG 521
Db 805 TCGCGGAGGAGCTGCTTGGACAAACACACAGCTGTGTGTTCCAGGACATGAACAG 864
QY 522 C-----GACCCGAGC-CATC-----C 537
Db 865 CCCCTTGCCCACTACTCTCATCTTCTCTCCCAACACACTATCTGACTGACTCCACATC 924
QY 538 AGGGGCAAGAGGAATACGTG---CTCTGT---GGGTCT---CCC-----CCACGCG- 582
Db 925 GGGGGGCC-AGCAGCAGAGGCTATGTAGGGCTTTGGCCAGGATGCGCGTGTCT 983
QY 583 -----CCT 585
Db 984 GAGCTGAGCTGTGGGAGGGCCAGGAGGGAGCCGCTCATCTATCATGAGCCATACCT 1043
QY 586 CCCC-----GGATCT-----GAGGAGAACAGACCGATCGCGCGCCA----- 623
Db 1044 CACCTCAAGATCTCTTTCGGGAGCTGGCCAG-CCGTGC-GCGACCATGCTTCAAG 1101
QY 624 CTGCGCCCTTA-ACTG-CATCAG--CCTGGG-----CTG-AGGCTGAGGA-CTGGCG 672
Db 1102 CTGTCCCTTACCCCTGCTATCTCTTCTGAGAACCACTGCGGGCTGAGAGCAGGCG- 1160
QY 673 AGGAGAGGCGCTCT-CTCTGCAC-----ACCTACTAG----- 705
Db 1161 -TGCCATGGCGCGCCACCTCTGCACCATCTCGGGGACATGCTGTGACACAGCGCTGG 1219
QY 706 -----TCACGAG--ACTTTAGG-----CGT-- 726
Db 1220 ACTCCCCAAATCCGAGAGGTGCCATCCCGAGAGAGCTGAAGGGCGCGGTCTCTGTGA 1279
QY 727 -GGGATT-----CCTACT-----CGTGTG-----TTCTAT----- 750
Db 1280 AGGGAAGAGAGTGGCGCTGCTCGGAGCAGAGTGGCGGCTCTGTGCGATCGGAGG 1339
QY 751 -----TTTTTGAAGAGACATTTAATAATGGT-----CA-----CGTT 787
Db 1340 AGGGGAGAGATACAGAGAGAGAGAGAGAGAGGAGGAGGCTGACGAGCAGAGCGCG 1399
QY 788 TGGTCTCTCTCAGATTTCTGAGGAAT---TGCTTTGTATTTATATTAATGATCACC 844
Db 1400 TG--GCCAAGCAGATCTCCCGGAGCTGTGCGGCTG--GCTGTGTACTGTCCACGCGCACCC 1456
QY 845 GACTGAGATATGTTT-----TACATATGTTCTG-TGGGGCTG-TTTTTT 889
Db 1457 GCCTCGGAGCCCTGACCCCTGCCCCCAACAGCCCAACCCCTGCCAGGTGAGCTCCCTCA 1516

QY 890 GTTA---TTAAACAAATAATTTAGATGTTGAAAAAAA 926
Db 1517 GCGAGCCCAAGCCCAAACTCAATTCGGAGCGCAGGAA 1556

RESULT 2

US-08-900-230-3
; Sequence 3, Application US/08900230
; Patent No. 6329197
; GENERAL INFORMATION:
; APPLICANT: Bard, Jonathan A.
; TITLE OF INVENTION: DNA ENCODING GALAN GALR3 RECEPTORS AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of The Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,230
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52241-C/JPW/ADM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-900-230-3

Query Match 44.4%; Score 2559; DB 4; Length 1417;
Best Local Similarity 43.0%; Pred.No. 3.1e-108;
Matches 564; Conservative 0; Mismatches 331; Indels 416; Gaps 86;

QY 1 GGGGC-----CCA--TTCTGTTTC---AGCCAGTC----- 25
Db 122 GGGGCTTTCCTGAGATTCCAGCTTCTTCCAGGTGCCCTGTGATGGGAGATGG 181
QY 26 -----GCCAAGAA-----TCATGAAAGTC-----GCCAGTGGCAGCACC- 59
Db 182 CTGATGCCAGAACATTTCACTGGACAGCCAGGGAGTGTGGGGCC-GTGGCAGTGCCT 240
QY 60 -----GCC-----ACCGC---CGCGCGGGCC----- 78
Db 241 GTGGTCTTTCCTTAATCTCTCTGGGCAAGTGGGCAATGGGCTGGTGGTGGCAGTG 300
QY 79 --CCAGTGC--GGCTGAAGGCC--GGCAAGA-----CAGC-----GAGCGGTGGC- 119
Db 301 CTCTGAGCCTGGC--CCGAGTGCCTGGCAGGAGCCTGGCAGCACCACCGACCTGTTTCA 359
QY 120 ----GGCGAGGTGTGGGTG---TCTG-----TCTGAGCAGAGCGTG----- 155
Db 360 CCTCAACTGCGGGTG-GCTGACCTCTGTTTCTCTGTGC---TGCGTGCCTTCCAGGC 416

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:17:00 ; Search time 23 Seconds
(without alignments)
2078.506 Million cell updates/sec

Title: X77956
Perfect score: 5766
Sequence: 1 GGGGCCCATCTGTTTCAGC.....TTAGATGGTGAAGAAAAA 926

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 11 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2596	45.0	2211	4	US-09-738-884-1
2	2559	44.4	1417	4	US-08-900-230-3
3	2431.5	42.2	1345	2	US-08-977-767-3
4	2412	41.8	2508	4	US-09-627-6508-7
5	2412	41.8	2508	4	US-09-436-063C-7
6	2397	41.6	2544	4	US-09-627-6508-3
7	2397	41.6	2544	4	US-09-436-063C-3
8	2385	41.4	2088	4	US-09-548-372D-13
9	2385	41.4	2088	4	US-09-548-367D-13
10	2385	41.4	2088	4	US-09-551-853D-13
11	2383.5	41.3	2601	4	US-09-436-063C-9
12	2383.5	41.3	2601	4	US-09-627-6508-9
13	2337.5	40.5	1400	4	US-08-640-915A-37
14	2327	40.4	1652	4	US-09-627-6508-1
15	2327	40.4	1652	4	US-09-436-063C-1
16	2314	40.1	1497	4	US-09-060-854B-2
17	2267.5	39.3	1917	4	US-09-627-6508-5
18	2267.5	39.3	1917	4	US-09-436-063C-5
19	2240	38.8	1128	4	US-09-627-6508-11
20	2240	38.8	1128	4	US-09-436-063C-11
21	1965	34.1	1388	4	US-09-463-048A-6
22	1772.5	30.7	908	4	US-08-714-741-44
23	1479.5	25.7	717	4	US-09-644-460-37
24	1355.5	23.5	605	4	US-09-976-594-616
25	1311	22.7	801	1	US-07-906-349A-6
26	1266.5	22.0	1184	4	US-10-153-064-89
27	1110	19.3	1332	1	US-07-609-716-41

28 1110 19.3 1332 3 US-08-475-411A-41 Sequence 41, Appl
29 1110 19.3 1332 3 US-08-478-029A-41 Sequence 41, Appl
30 942 16.3 1136 4 US-08-806-029-9 Sequence 9, Appl
31 942 16.3 1177 1 US-07-609-716-31 Sequence 31, Appl
32 942 16.3 1177 1 US-08-175-155-29 Sequence 29, Appl
33 942 16.3 1177 1 US-08-477-509B-64 Sequence 35, Appl
34 942 16.3 1177 2 US-08-707-237A-35 Sequence 64, Appl
35 942 16.3 1177 3 US-08-482-085B-64 Sequence 31, Appl
36 942 16.3 1177 3 US-08-475-411A-31 Sequence 31, Appl
37 942 16.3 1177 3 US-08-478-029A-31 Sequence 31, Appl
38 942 16.3 1177 4 US-09-444-791A-64 Sequence 64, Appl
39 905.5 15.7 1059 1 US-08-175-155-48 Sequence 48, Appl
40 905.5 15.7 1059 2 US-08-707-237A-54 Sequence 54, Appl
41 905.5 15.7 1059 4 US-08-806-029-10 Sequence 10, Appl
42 905.5 15.7 1101 1 US-08-477-509B-83 Sequence 83, Appl
43 905.5 15.7 1101 3 US-08-482-085B-83 Sequence 83, Appl
44 905.5 15.7 1101 4 US-09-444-791A-83 Sequence 83, Appl
45 902.5 15.7 1038 1 US-07-609-716-36 Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-738-884-1
; Sequence 1, Application US/09738884
; Patent No. 6391606

; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000849
; CURRENT APPLICATION NUMBER: US/09/738,884
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1
; LENGTH: 2211
; TYPE: PRT
; ORGANISM: Human
US-09-738-884-1

Query Match 45.0%; Score 2596; DB 4; Length 2211;
Best Local Similarity 38.8%; Pred. No. 9.3e-110;
Matches 597; Conservative 0; Mismatches 305; Indels 638; Gaps 89;

QY 1 GGGGCCCATCTGTTTCAGCCAG-----TCCG-----CAAGATC----- 35
Db 41 GGGGCTC---CCGGCTCCGCAAGATCCGCTCGGCACAGTGCGCAGAGGGCTGTACC 97
QY 36 -----ATGAAAGTC-CCAGTG-----GCAGCAC-----CGCCACCGC 67
Db 98 GGCTGCAGGAGGACGGCTGAGCGTGTGTTCCAGCGGGGCATCCCGGTGGCCATCGC 157
QY 68 CGC-----CGCGGGCC-CCAG----- 82
Db 158 AGCAGATCTTCTTCGTGTCAGCAGCATCGAGGGGTCCGCGAGGGCCACCATCGAGGGCC 217
QY 83 -----CTCGCGCTG-----AAGGCC 98
Db 218 TGGCGCGCTTCGGGGGTGCTTCGCGCCAGCGGCTCCCTCACCATCGCTTCAAGGGCC 277
QY 99 G--GCAGAG--CAG---CGAGCGGTG---CGGGCGAGGT---GGTGGCT 135
Db 278 GCGCGAGAGACCTTGAGACCTTGCGGCCGCCACCGCTGAGGAGCGGAGCGCTGGTGGCG 337
QY 136 GTCTGTGTAGAGAGAGCGTG-----GCCATTCGCGCTGC-----CG 172
Db 338 GTCTGCAAGACTCCCGCGCGCTGAGCGCCATGAGCCAGCGCGGCTAGACCACT 397
QY 173 GG-----GCGCGCGGGG--CGCTGCTGCTGAGCGAGCAGC--AGG 216

Search completed: May 7, 2004, 15:17:53
Job time : 85 secs

```

PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 02-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 04-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 10-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 26-AUG-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 10-SEP-1998; 98US-0100388P.
PR 15-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-01019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100864P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 24-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.

PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-00168978.

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred. No. 3.5e-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

QY 1 GGGGCCCATTTCTGTTTTCAGC-----CAGTCGCCAAG-----AATCATGAAGTCGCCAGTG 51
DB 40 GTGGCAGCGCGGTAGCAGCAATGGACTTTCTCTGGGAAACCGTTACGCTCTCCAGTG 99
QY 52 G--CAGCAC--CGCCACCGCGCGCGGCG--CCAGCTCGGGCTGA--AGGCC-----98
DB 100 GGACGCGCATCGAGAAAGCCAGATGGTCCCTCAGAGCGAGACTGGGCGCTCAAC 159
QY 99 --CG-----CAAGACAG-----CGAGCG 114
DB 160 ATGAGATCTCGACATCATCAAGAGAGCGAGGAGGTCCCAAGATGCCCTCCGAGCA 219
QY 115 GT-----CGGG-----CGAGTGGTGC--CCTGT--CTGTCT 142
DB 220 GTAAAGAGAGATCTGTGGGAATAAGAACTTCCACGAGGTGATGCTGCTCTCAGATC 279
QY 143 TGAGCA-----GAGCGTGGCCATCTCGCGCT-----GCCGGGCG 176
DB 280 TTAGAACCTGTGTCAAGAACTCGCGGCGACCGCTTCCACGTGCTGTGGTCCAGCAGGAC 339
QY 177 GCGGGGCGCG-----CCTGCTGCCCTGCTGACGAGCAGCAGGTAACG 222
DB 340 TTCTGGAGAGTGTGCTGTGAGGACCATCTGCGC--AAGAACAACCCACCCACCATCG 397
QY 223 TGC-----TGCTTACGACAT--GAAC--GGCT-----GTTACTCA--CGCC--258
DB 398 TGCATGACAAAGTGTCAACCTCATCCAGTCTCTGGGCTGACGCGTTCGCGAGTCCGCGG 457
QY 259 -TC---AAGGAGTGTGTGCCA-----CCCTGC-----CCG--AGAACCGC 293
DB 458 ATCTGACAGGTG--TGGTCAACATCTATGAGGACCTCGGAGGAAGGCGCTGGAGTTCCCC 516
QY 294 AAGG-----TGAGCAAGTGGAGATCTCCA-----GCACTGTCATCGACT---333
DB 517 ATGACTGACCTGGACATGTGTCAACATCCACACCCAGAGGACCGTGTCACTCAG 576
QY 334 ---ACATCAGGAC---CTTCAGTTGGAGTGAATCGGAATCGGAATGGGACCCCG 386
DB 577 AGACACAATCAGCAGCAGGATTTCTGTGGCACTG-CTCT--CAGCCAGCAAGAGGACTCT- 632
QY 387 GGGGGC--GAGGCTCGCGTCCGGCTCGGCTCAGCAC--CCTCAACGG--CGAGATC 440
DB 633 ---GGCCAGCATGTCGCCCTCTG---CCGCCCCCGCCATCTCTCCGTTGACACGCC 686
QY 441 AGCGCCCTGACGGCCGAGCGCGCATGCTTCTCTGCG-----GACGATCG-----485
DB 687 ATAGACCAAC--CCCGAACAGATTGGGAAGCTGCGCAGTGAGCTGGAGATGTTAGTGG 745
QY 486 -ATCTTG-----TGTC-----GCTGA--AGC---GCTTCCCCCAG--GGACCGG 521
DB 746 GAACGTGAGGCTGATGTCGAGATGCTGACGGAGTGTGTGCC--CACCCAGCGCGAGCCCG 804
QY 522 CGGACC-----CC-AGCCATCCAG-----GGGGCA 545
DB 805 CAGACCTGGAGCTGCTCAGGAGCTCAACCGCAGCTGCGGAGCCATGACGACGCGGTCC 864
QY 546 AGAGGAATTACGTGCTGTGTGGTCT--TCCCCAAAGCGCC--TC---GCCGATCTGAG 598
DB 865 TGAGTGATACCTGCTC--CGGGCCCATGCCCCAAGAGGCGCTTACAGAGCCACATGCC 922
QY 599 GGAGAACAAAGACCGATC---GGC--GGCCACTG-----CG-----CCC 631
DB 923 AG---TCGAGGCTGCTGGAGGCTGGCCACAGATGGGAATTTCTCCGAGCGCTATTGCC 979

```



```

QY      672  --GAGGAGAGGGCGCTCTCTCTGACACCTACTAGTC--ACCAGAGACTTTAGGGGTG 727
Db      1037  TGGAGGTGGAGGCCCC---CTGAGGAGCTGC---GGCGGCCAGGTAC---GAAAGCTG 1087
QY      728  GGAATCCATCG-TGTG-----TTTCTATTTTGAAGCAGACATTTAAATAATGTT 781
Db      1088  CAACTGCGCGCAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC 1128
QY      782  CACGTTTGTGTTCTTCAGATTCTGAGG-----AAATTGCTTTGTATTGTA 828
Db      1129  TCGAGGTGAGGCTTCAGGGAGTCTGGGGCCCACTGCCCTCGCTGCTTGGCCCTCCA 1188
QY      829  T-ATTACAATGATC-----ACCG-----ACTGAG-----AATATGTTT--TA 863
Db      1189  TCCTTCTCTGTCTCTTCTTGGCGCGGCACACAGACTGGGCTCACTCTTGGTGTATC 1248
QY      864  CAAATGTTCTGTGGGCTGT-TTTTGTGTTATTAACAATTAATTAAGTGTGAAAAA 921
Db      1249  CTTCTTACTG-GGAGAGTGCTTTTGT--ATCCCAATTA-----AAGGTAGAAAA 1298

RESULT 15
ABU88959
ID      ABU88959 standard; protein; 1300 AA.
XX
XX
AC      ABU88959;
XX
XX
DT      09-JUL-2003 (first entry)
XX
DE      Novel human secreted and transmembrane PRO protein #2.
XX
KW      Human; gene therapy; chondrocyte stimulation; TNF-alpha release;
KW      chondrocyte proliferation; chondrocyte differentiation; tumour detection;
KW      tissue typing.
XX
OS      Homo sapiens.
XX
PN      US2003036133-A1.
XX
PD      20-FEB-2003.
XX
PF      27-JUN-2002; 2002UG-00184630.
XX
PR      18-SEP-1997; 97US-0059263P.
PR      18-SEP-1997; 97US-0059266P.
PR      17-OCT-1997; 97US-0062250P.
PR      21-OCT-1997; 97US-0063486P.
PR      24-OCT-1997; 97US-0063120P.
PR      24-OCT-1997; 97US-0063121P.
PR      28-OCT-1997; 97US-0063540P.
PR      28-OCT-1997; 97US-0063541P.
PR      28-OCT-1997; 97US-0063544P.
PR      29-OCT-1997; 97US-0063734P.
PR      31-OCT-1997; 97US-0063870P.
PR      31-OCT-1997; 97US-0064103P.
PR      13-NOV-1997; 97US-0065311P.
PR      21-NOV-1997; 97US-0066120P.
PR      24-NOV-1997; 97US-0066466P.
PR      11-DEC-1997; 97US-0066772P.
PR      12-DEC-1997; 97US-0069435P.
PR      17-DEC-1997; 97US-0069870P.
PR      18-DEC-1997; 97US-0068017P.
PR      10-MAR-1998; 98US-0077450P.
PR      11-MAR-1998; 98US-0077632P.
PR      11-MAR-1998; 98US-0077649P.
PR      20-MAR-1998; 98US-0078886P.
PR      20-MAR-1998; 98US-0078939P.
PR      27-MAR-1998; 98US-0079664P.
PR      27-MAR-1998; 98US-0079786P.
PR      31-MAR-1998; 98US-0080107P.
PR      31-MAR-1998; 98US-0080194P.

```

```

PR      01-APR-1998; 98US-0080327P.
PR      01-APR-1998; 98US-0080333P.
PR      08-APR-1998; 98US-0081049P.
PR      08-APR-1998; 98US-0081070P.
PR      09-APR-1998; 98US-0081195P.
PR      15-APR-1998; 98US-0081838P.
PR      21-APR-1998; 98US-0082568P.
PR      21-APR-1998; 98US-0082569P.
PR      22-APR-1998; 98US-0082704P.
PR      22-APR-1998; 98US-0082797P.
PR      22-APR-1998; 98US-0083322P.
PR      29-APR-1998; 98US-0083495P.
PR      29-APR-1998; 98US-0083496P.
PR      29-APR-1998; 98US-0083499P.
PR      29-APR-1998; 98US-0083559P.
PR      05-MAY-1998; 98US-0084366P.
PR      06-MAY-1998; 98US-0084414P.
PR      07-MAY-1998; 98US-0084639P.
PR      07-MAY-1998; 98US-0084640P.
PR      07-MAY-1998; 98US-0084643P.
PR      15-MAY-1998; 98US-0085579P.
PR      15-MAY-1998; 98US-0085580P.
PR      15-MAY-1998; 98US-0085582P.
PR      15-MAY-1998; 98US-0085700P.
PR      18-MAY-1998; 98US-0086023P.
PR      22-MAY-1998; 98US-0086392P.
PR      22-MAY-1998; 98US-0086486P.
PR      28-MAY-1998; 98US-0087098P.
PR      28-MAY-1998; 98US-0087208P.
PR      02-JUN-1998; 98US-0087609P.
PR      02-JUN-1998; 98US-0087759P.
PR      03-JUN-1998; 98US-0087827P.
PR      04-JUN-1998; 98US-0088025P.
PR      04-JUN-1998; 98US-0088028P.
PR      04-JUN-1998; 98US-0088029P.
PR      04-JUN-1998; 98US-0088033P.
PR      04-JUN-1998; 98US-0088326P.
PR      05-JUN-1998; 98US-0088167P.
PR      05-JUN-1998; 98US-0088202P.
PR      05-JUN-1998; 98US-0088212P.
PR      05-JUN-1998; 98US-0088217P.
PR      09-JUN-1998; 98US-0088655P.
PR      10-JUN-1998; 98US-0088722P.
PR      10-JUN-1998; 98US-0088738P.
PR      10-JUN-1998; 98US-0088740P.
PR      10-JUN-1998; 98US-0088811P.
PR      10-JUN-1998; 98US-0088824P.
PR      10-JUN-1998; 98US-0088825P.
PR      11-JUN-1998; 98US-0088826P.
PR      11-JUN-1998; 98US-0088861P.
PR      11-JUN-1998; 98US-0088863P.
PR      11-JUN-1998; 98US-0088876P.
PR      12-JUN-1998; 98US-0089090P.
PR      12-JUN-1998; 98US-0089105P.
PR      16-JUN-1998; 98US-0089512P.
PR      16-JUN-1998; 98US-0089514P.
PR      17-JUN-1998; 98US-0089538P.
PR      17-JUN-1998; 98US-0089598P.
PR      17-JUN-1998; 98US-0089653P.
PR      18-JUN-1998; 98US-0089908P.
PR      19-JUN-1998; 98US-0089952P.
PR      22-JUN-1998; 98US-0090246P.
PR      22-JUN-1998; 98US-0090252P.
PR      22-JUN-1998; 98US-0090254P.
PR      24-JUN-1998; 98US-0090439P.
PR      24-JUN-1998; 98US-0090435P.
PR      24-JUN-1998; 98US-0090444P.
PR      24-JUN-1998; 98US-0090461P.
PR      24-JUN-1998; 98US-0090535P.
PR      24-JUN-1998; 98US-0090540P.
PR      25-JUN-1998; 98US-0090676P.
PR      25-JUN-1998; 98US-0090678P.
PR      25-JUN-1998; 98US-0090688P.

```

```
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 24-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 18-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097922P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 01-SEP-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 02-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 09-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-0101751P.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102457P.
PR 30-SEP-1998; 98US-0102570P.
PR 01-OCT-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 02-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102985P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-00168978.
PR 07-OCT-1998; 98US-0103395P.

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred. No. 3 Se-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

Qy 1 GGGGCCCATCTCTTTTCAGC-----CAGTCGCCAAG-----AATCATGAAGTGGCCAGTG 51
Db 40 GTGGCAGCGCGGTAGCAGCAATGGACTTTCTCTCTGGGAACCCGTTACGCTCTCCAGTG 99
Qy 52 G--CAGCAC--CGCCACCGCGCGCGGCG--CCAGCTCGCGCTGA--AGGCC-----98
Db 100 GGACAGGCATCGAGAAGCCACAGATGCTCCCTCGAGAGCGAGGACTGGGCCCTCAAC 159
Qy 99 --GG-----CAAGACAG-----CGAGG 114
Db 160 ATGGAGATCTGCACATCATCAACGAGCGAGGAGGTCCCAAAGATGCCCTCCGAGCA 219
Qy 115 GT-----GCGGG-----CGAGTGGTGC--GCTGT--CTGTC 142
Db 220 GTAAAGAGAGATCGTGGGGAATAGAACTTCCACGAGGTGATGCTGCTCTCACATGTC 279
Qy 143 TGAGCA-----GAGCGTGCCATCTCGGCT-----GCCGGGG 176
Db 280 TTAGAAACCTGTCTCAAGAACTGCGGGCACCCCTTCCAGCTGCTGGTGCCAGCGAGGAC 339
Qy 177 GCGGGGCGCG-----CCTGCTCCCTGCTGTGGAGCACTCTCTGCC--AAGAACACCCACCCATCG 397
Db 340 TTCGTGGAGAGTGTGCTGTGGAGCACTCTCTGCC--AAGAACACCCACCCATCG 397
Qy 223 TGC-----TGCTCTAGACAT---GAAC--GGCT-----GTTACTCA---CGCC--258
Db 398 TGCATGACAAAGTGTCTCAACCTCATCCAGTCTGTGGCTGAGCGCTTCCGAGCTCGCCG 457
Qy 259 -TC---AAGGAGTGTGTGCGCA-----CCCTGC-----CCC--AGAACCG 293
Db 458 ATCTGACAGTG--TGGTCAACCTCATCCAGTCTGTGGCTGAGCGCTTCCGAGCTCGCCG 516
Qy 294 AAG-----TGAGCAAGTGGAGATTCTCCA-----GCAGTCTAGCT---333
Db 517 ATGACTGACCTGGACATGCTGCACCCATCCACACACCCAGAGGACCGTGTTCACATCAG 576
Qy 334 ---ACATCAGGAC---CTTCAGTTGGAGCTGAATCGGAATCGGAATCGGAACCGCC 386
Db 577 AGACACATCAGGACAGGATTCTGTGGCACTG-ATC--CAGCCAGCAAGAGGACTCT-632
Qy 387 GGGGGC--GAGGCTCGCGTCCGGCTCGGCTCAGCAG--CCTCAACGG---CGAGTC 440
Db 633 ---GGCCAGCATGCTGCGCTCTG---CCGCGCGCGCCATCTCTCGGTGACACGCGCC 686
Qy 441 AGCGCCCTGACCGCCGAGGCGGCGATGCTTCTGCG-----GACGATCGC-----485
Db 687 ATAGCACCAAC-CCCGGAACAGATTGGGAAGCTGCGCAGTGAAGTGGTGAAGTGG 745
Qy 486 -ATCTTG-----TGTC-----GCTGA--AGC---GCCTCCCGCAG--GGACCG 521
Db 746 GAACGTGAGGGTATGTCGAGATGCTGACGAGCTGTGTC--CACCAGGCGGAGCGCC 804
Qy 522 CGGACC-----CC-AGCCATCCAG-----GGGGCA 545
Db 805 CAGACTGGAGTGTGTCAGGAGCTCAACCGCACGTGCGGAGCCATGACAGCGGCTCC 864
Qy 546 AGAGGATTACGTGCTGTGGTC--TCCCGCAAGCGCC--TC---GCCGATCTGAG 598
Db 865 TGAGTGATACCTGTGTC--CGGGCCCATGCCCCAAGGAGCCCTTCAGAGCCACACTGCC 922
Qy 599 GGAGAACAAAGACCGATC---GCG--GGCCACTG-----CG-----CCC 631
Db 923 AG--TCGAGGCTGGCTGGAGCTGGCCACAGTGGAAATTCGCGGAGCTATTGTGCC 979
Qy 632 TTAAGTGCATC--CAGCTGGGG--CTGAGGCTGAGGACTGGC-----671
Db 980 TACCCTGC-TCTGCTCATGGGGCCCTCATGGCTTTGG--CTGGCCACTGAGGTTAGGTT 1036
```

QY	728	GGATTCCACTCG-TGTG-----TTTCTATTTTGTAAAGACACACATTTTAAAAATGCT	781
DB	1088	CAACTCTCGCGCAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC	1128
QY	782	CACGTTTGGTCTTCTCAGATTCTGAGG-----AAATGCTTTGTATGTA	828
DB	1129	TCAGGTGAGGTTTCAGGGGATGCTGGGCCCACTGCCCTCGGCTGCTTGCCTCCA	1188
QY	829	T-ATTACATGATC-----ACCG-----ACTGAG-----AAATGCTTT--TA	863
DB	1189	TCCTTCCCTGCTTCTTCTGCGCGGACACACACTGGGCTCCTCTTGTGATC	1248
QY	864	CAATAGTTCCTGGGGCTGT-TTTTTGTATTAAACAATAATTAGATGGTGA	921
DB	1249	CTCTTGACTG-GGAGAGGTGCTTTTGT--ATCCCAATTA-----AAGGTAGAAAA	1298

RESULT 14

ABO00408

ID ABO00408 standard; protein; 1300 AA.

XX AC ABO00408;

XX DT 06-AUG-2003 (first entry)

XX DE Novel human secreted and transmembrane PRO protein #2.

XX DE

XX KW Human; gene therapy; tumour necrosis factor alpha; TNF-alpha;

XX KW chondrocyte stimulation; tumour; tissue typing.

XX OS Homo sapiens.

XX PN US2003032101-A1.

XX PD 13-FEB-2003.

XX PF 17-JUN-2002; 2002US-00173695.

XX PR 18-SEP-1997; 97US-0059263P.

PR 18-SEP-1997; 97US-0059266P.

PR 17-OCT-1997; 97US-0062250P.

PR 21-OCT-1997; 97US-0063485P.

PR 24-OCT-1997; 97US-0063120P.

PR 24-OCT-1997; 97US-0063121P.

PR 28-OCT-1997; 97US-0063540P.

PR 28-OCT-1997; 97US-0063541P.

PR 28-OCT-1997; 97US-0063544P.

PR 29-OCT-1997; 97US-0063734P.

PR 31-OCT-1997; 97US-0063870P.

PR 31-OCT-1997; 97US-0064103P.

PR 13-NOV-1997; 97US-0065311P.

PR 21-NOV-1997; 97US-0066120P.

PR 24-NOV-1997; 97US-0066466P.

PR 11-DEC-1997; 97US-0066772P.

PR 12-DEC-1997; 97US-0069335P.

PR 18-DEC-1997; 97US-0069425P.

PR 18-DEC-1997; 97US-0069870P.

PR 10-MAR-1998; 98US-0077450P.

PR 11-MAR-1998; 98US-0077632P.

PR 11-MAR-1998; 98US-0077649P.

PR 20-MAR-1998; 98US-0078886P.

PR 20-MAR-1998; 98US-0078939P.

PR 27-MAR-1998; 98US-0079664P.

PR 27-MAR-1998; 98US-0079786P.

PR 31-MAR-1998; 98US-0080107P.

PR 31-MAR-1998; 98US-0080194P.

PR 01-APR-1998; 98US-0080327P.

PR 01-APR-1998; 98US-0080333P.

PR 08-APR-1998; 98US-0081049P.

PR 08-APR-1998; 98US-0081070P.

PR 09-APR-1998; 98US-0081195P.

PR 15-APR-1998; 98US-0081838P.

PR 21-APR-1998; 98US-0082568P.

PR 21-APR-1998; 98US-0082569P.

PR 22-APR-1998; 98US-0082704P.

PR 22-APR-1998; 98US-0082797P.

PR 28-APR-1998; 98US-0083322P.

PR 29-APR-1998; 98US-0083495P.

PR 29-APR-1998; 98US-0083496P.

PR 29-APR-1998; 98US-0083499P.

PR 29-APR-1998; 98US-0083559P.

PR 05-MAY-1998; 98US-0084366P.

PR 06-MAY-1998; 98US-0084414P.

PR 07-MAY-1998; 98US-0084639P.

PR 07-MAY-1998; 98US-0084643P.

PR 15-MAY-1998; 98US-0085579P.

PR 15-MAY-1998; 98US-0085580P.

PR 15-MAY-1998; 98US-0085582P.

PR 15-MAY-1998; 98US-0085700P.

PR 18-MAY-1998; 98US-0086023P.

PR 22-MAY-1998; 98US-0086392P.

PR 22-MAY-1998; 98US-0086486P.

PR 28-MAY-1998; 98US-0087098P.

PR 28-MAY-1998; 98US-0087208P.

PR 02-JUN-1998; 98US-0087609P.

PR 02-JUN-1998; 98US-0087759P.

PR 03-JUN-1998; 98US-0087827P.

PR 04-JUN-1998; 98US-0088025P.

PR 04-JUN-1998; 98US-0088028P.

PR 04-JUN-1998; 98US-0088033P.

PR 04-JUN-1998; 98US-0088326P.

PR 05-JUN-1998; 98US-0088167P.

PR 05-JUN-1998; 98US-0088202P.

PR 05-JUN-1998; 98US-0088212P.

PR 05-JUN-1998; 98US-0088217P.

PR 09-JUN-1998; 98US-0088655P.

PR 10-JUN-1998; 98US-0088722P.

PR 10-JUN-1998; 98US-0088738P.

PR 10-JUN-1998; 98US-0088740P.

PR 10-JUN-1998; 98US-0088811P.

PR 10-JUN-1998; 98US-0088824P.

PR 10-JUN-1998; 98US-0088825P.

PR 10-JUN-1998; 98US-0088826P.

PR 11-JUN-1998; 98US-0088861P.

PR 11-JUN-1998; 98US-0088863P.

PR 11-JUN-1998; 98US-0088876P.

PR 12-JUN-1998; 98US-0089090P.

PR 12-JUN-1998; 98US-0089105P.

PR 16-JUN-1998; 98US-0089512P.

PR 16-JUN-1998; 98US-0089514P.

PR 17-JUN-1998; 98US-0089538P.

PR 17-JUN-1998; 98US-0089598P.

PR 17-JUN-1998; 98US-0089653P.

PR 18-JUN-1998; 98US-0089908P.

PR 19-JUN-1998; 98US-0089952P.

PR 22-JUN-1998; 98US-0090246P.

PR 22-JUN-1998; 98US-0090252P.

PR 22-JUN-1998; 98US-0090254P.

PR 24-JUN-1998; 98US-0090429P.

PR 24-JUN-1998; 98US-0090435P.

PR 24-JUN-1998; 98US-0090444P.

PR 24-JUN-1998; 98US-0090461P.

PR 24-JUN-1998; 98US-0090535P.

PR 24-JUN-1998; 98US-0090540P.

PR 25-JUN-1998; 98US-0090676P.

PR 25-JUN-1998; 98US-0090678P.

PR 25-JUN-1998; 98US-0090688P.

PR 25-JUN-1998; 98US-0090690P.

PR 25-JUN-1998; 98US-0090694P.

PR 25-JUN-1998; 98US-0090695P.

PR 25-JUN-1998; 98US-0090696P.

PR 26-JUN-1998; 98US-00105413.

PR	01-JUL-1998;	98US-0091359P.	QY	1	GGGGCCCATTTCTGTTTTCAGC-----CAGTCGCCAAG---AATCATGAAGTCGCCAGTG	51
PR	01-JUL-1998;	98US-0091544P.	Db	40	GTGGCAGCGCGGTAGCAGCAATGGACTTCTCTGGGAACCCGTTCACTCTCCAGTG	99
PR	02-JUL-1998;	98US-0091478P.	QY	52	G--CAGCAC--CGCCACCGCCCGCGGC--CCAGACTGGCGCTGA--AGGCC-----	98
PR	02-JUL-1998;	98US-0091628P.	Db	100	GGACAGCGCATCGAAGAGCCACATGGCTCCCTGCAGAGCAGACTGGGCCCTCAAC	159
PR	04-AUG-1998;	98US-0095282P.	QY	99	--GG-----CAAGACAG-----CGAGCG	114
PR	10-AUG-1998;	98US-0095998P.	Db	160	ATGGAGATCTGGACATCATCAAGAGCGGAGGAGTCCCAAAGATGCCCTCCGAGCA	219
PR	17-AUG-1998;	98US-0096757P.	QY	115	GT-----GCGGG-----CGAGTGGTGC--GCTGT--CTGTC	142
PR	17-AUG-1998;	98US-0096867P.	Db	220	GTAAGAAGAGAAATCGTGGGAATAAGAACTTCCACGAGGTGATGCTGCTCTCACAGTC	279
PR	16-AUG-1998;	98US-0096949P.	QY	143	TGAGCA-----GAGCGTGCCCATCTCGCGCT-----GCCGGGCG	176
PR	18-AUG-1998;	98US-0095852P.	Db	280	TTAGAAACCTGTGTCAAGAACTCGCGGCACCGCTTCCACGTGTGTGTCGCGCAGGAC	339
PR	26-AUG-1998;	98US-0097954P.	QY	177	GCGGGGCGCG-----CCTGCTGCCCTGCTGGACGAGCAGCAGGTAAAGC	222
PR	26-AUG-1998;	98US-0097955P.	Db	340	TTCTGGAGAGTGTGTGTGGAGACCATCTCTGCC--AAGAAACAACCCACCCACCATCG	397
PR	26-AUG-1998;	98US-0097971P.	QY	223	TGC-----TGCTCTACGACAT--GAAC--GGCT-----GTTACTCA--CGCC--	258
PR	26-AUG-1998;	98US-0097974P.	Db	398	TGATGACAAAGTGTCTCAACTCATCCAGTCTGTGGCTGACGCTTCCGACGTGCCCCG	457
PR	26-AUG-1998;	98US-0098014P.	QY	259	-TC---AAGAGCTGTGTGCCA-----CCCTGC-----CCC--AGAACCGC	293
PR	01-SEP-1998;	98US-0098716P.	Db	458	ATCTGACAGTG--TGCTCACCATCTATGAGGACCTCGGAGGAAGGCCCTGGAGTTCCTCC	516
PR	02-SEP-1998;	98US-0098803P.	QY	294	AAGG-----TGAGCAAGGTGGAGATTCTCCA-----GCACGTCTCATGACT---	333
PR	02-SEP-1998;	98US-0098821P.	Db	517	ATGACTGACCTGGACATGTGTACCCATCCACACCCAGAGGACCGTGTTCACATCAG	576
PR	16-SEP-1998;	98US-0100662P.	QY	334	----ACATCAGGAC---CTTCAGTTGGAGTGAATCGGAATCGGAATGGGACCCCC	386
PR	16-SEP-1998;	98US-0100664P.	Db	577	AGACACAATCAGGACAGGATTCTGTGGCACTG-ACTC--CAGCCAGCAAGAGGACTCT-	632
PR	17-SEP-1998;	98US-0100683P.	QY	387	GGGGGC--GAGGCTCGCGTCCGGCTCAGCAGC--CCTCAACCG--CGAGATC	440
PR	17-SEP-1998;	98US-0100919P.	Db	633	---GGCCAGCATGCTGCCCTCTG---CCGCCGCCCCCTACTCTCCGGTGACAGCCC	686
PR	18-SEP-1998;	98US-0100849P.	QY	441	AGGCGCCTGACGGCCGAGGCGGCATGCTTCTGTGG--GACCATGCG-----	485
PR	23-SEP-1998;	98US-0101471P.	Db	687	ATAGCACCAAC--CCCGAACAGATTGGGAAGCTGCGCAGTGAGCTGGAGATGTTGAGTG	745
PR	23-SEP-1998;	98US-0101472P.	QY	486	-ATCTTG-----TGTC-----GCTGA---AGC---GCCTCCCCCAG--GGACCGG	521
PR	23-SEP-1998;	98US-0101477P.	Db	746	GAACGTGAGGGTGTATGTCCGAGATGCTGACGGAGTGGTGCC--CACCCAGGCCAGCCCC	804
PR	24-SEP-1998;	98US-0101738P.	QY	522	CGGACC-----CC-AGCCATCCAG-----GGGGCA	545
PR	24-SEP-1998;	98US-0101743P.	Db	805	CAGACCTGGAGCTGCTGCAGGAGCTCAACCGCAGCTGCCGAGCCATGCAGCAGCGGTCC	864
PR	29-SEP-1998;	98US-0102207P.	QY	546	AGAGAAATTACGTCTGTGTGGTGC--TCCCCAAAGCGCC--TC---GCCGATCTGAG	598
PR	29-SEP-1998;	98US-0102240P.	Db	865	TGAGTGATACCTGCTC--CGGGCCCATGCCCCAAGAGGCCCTTCAGAGCCCCACATGCC	922
PR	30-SEP-1998;	98US-0102487P.	QY	599	GGAGAACAAGACCGATC---GGC-GGCCACTG-----CG-----CCC	631
PR	30-SEP-1998;	98US-0102570P.	Db	923	AG---TCGAGGCTGGCTGAGGCTGCCACATGGAAATTCGCCGAGCCTATTGTCCC	979
PR	01-OCT-1998;	98US-0102684P.	QY	632	TTAACTGCATC--CAGCTGGG--CTGAGGCTGAGCACTGGC-----	671
PR	02-OCT-1998;	98US-0102687P.	Db	980	TACCTGCTGCTGTGATGGGGCCCCCATGGCTTGG--CTGGCCACTGAGGTAGGGTG	1036
PR	06-OCT-1998;	98US-0103258P.	QY	672	--GAGGAGGGCGCTCTCTGACACCTACTAGTC--ACCAGAGACTTTTAGGGGTG	727
PR	07-OCT-1998;	98US-00168978.	Db	1037	TGGAGGTGTGAGGCCCC---CTGAGGAGTGC--GGCGGCCCGAGGTAC---GAAGCTG	1087

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred. No. 3.5e-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

QY 782 CACGTTTGGTCTCTCAGATTCTGAGG-----AAATTGCTTTGTATTGTA 828
Db 1129 TCAGGTGAGGCTTACGGGATGCTGGGGCCCACTGCGCTCCCTTGGCCCTCCA 1188
QY 829 T-ATTACAATGATC-----ACCG-----ACTGAG-----AAATATGTTT--TA 863
Db 1189 TCCTTCCTCTGTTCTCTCTTCGCGGGGACACACAGCACTGGGCTCACCTCTTGGTTGATC 1248
QY 864 CAATAGTTCTGGGGGCTCT-TTTTGTGTTATTAAACAATAATTAGATGTTGAAAA 921
Db 1249 CTCCTGTACTG-GGAGAGTGCTTTTGT--ATCCCAATTA-----AAGGTAGAAAA 1298

RESULT 13
ABU85269
ID ABU85269 standard; protein; 1300 AA.
XX
AC ABU85269;
XX
DT 30-JUN-2003 (first entry)
XX
DE Novel human secreted and transmembrane protein #2.
XX
KW Human; secreted and transmembrane protein; PRO; cytosolic; gene therapy;
KW chondrocyte stimulator; chromosome mapping; gene mapping;
KW transgenic animal; knock-out animal; tumour.
XX
OS Homo sapiens.
XX
PN US2003032114-A1.
XX
XX
PD 13-FEB-2003.
XX
PF 20-JUN-2002; 2002US-00176919.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 28-OCT-1997; 97US-0063734P.
PR 29-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077649P.
PR 11-MAR-1998; 98US-0078866P.
PR 20-MAR-1998; 98US-0078939P.
PR 20-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 28-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083599P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-008876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-008953P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.

QY 829 T-ATTACATGATC-----ACCG-----ACTGAG-----AATATTGTTT--TA 863
Db 1189 TCCTTCTCTGTTCTTCTGCGCGGACACAGAGCTGGGGCTACCTCTTGGTGTATC 1248
QY 864 CAATAGTCTCTGGGGTGT-TTTTGTGTTATTAAACAATAATTAGATGTTGAAAA 921
Db 1249 CTCCTGTACTG-CGAGAGGTGCTTTTGT--ATCCCAATTA-----AAGGTAGAAAA 1298

RESULT 12
ABU91965
ID ABU91965 standard; protein; 1300 AA.
XX AC ABU91965;
XX DT 11-AUG-2003 (first entry)
XX DE Novel human secreted and transmembrane PRO protein #2.
XX KW Human; gene therapy; chromosome identification; tissue typing.
XX OS Homo sapiens.
XX PN US2003027277-A1.
XX PD 06-FEB-2003.
XX PF 21-JUN-2002; 2002US-00176985.
XX PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079766P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080154P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 02-JUN-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 03-JUN-1998; 98US-0087759P.
PR 04-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 05-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 09-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 11-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 24-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.

```
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 17-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096881P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0098843P.
PR 10-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 15-SEP-1998; 98US-0100652P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-0101930P.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 30-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103259P.
PR 06-OCT-1998; 98US-0103449P.

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred. No. 3.5e-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

QY 1 GGGGCCCATCTGTTTCACG-----CAGTGGCCAAG---AATCATGAAGTGGCCAGTGT 51
DB 40 GTGGCAGCGGGGTAGCAGCAATGAGATTTCTCTGGGGAACCCGTTTCACTTCAGTGT 99
QY 52 G--CAGCAC--CGCCACCGCCCGCGGC--CCGAGCTGCGCCTGA--AGGCC----- 98
DB 100 GGACAGCGCATCGAAGAACACAGATGGCTCCCTCGAGAGCAGGACTGGGCCCTCAAC 159

99 --CG-----CAGACAG-----CGAGCG 114
160 ATGAGATCTGCGACATCATCAACGAGACGAGGAGGTCCCAAAGATGCCCTCCGAGCA 219
115 GT-----CGCGG-----CGAGGTGTCG--GCTGT--CTGTC 142
220 GTAAAGAGAGATCTGCGGAATAGAACTTCCAGAGGTGATGCTGCTCTCAAGTTC 279
143 TGAGCA-----GAGCGTGCCCATCTGCGCT-----GCCGGGCG 176
280 TTAGAAACCTGTGTCAAGAACTGCGGCGCACCGCTTCCAGCTGCTGCTGGCCAGCAG 339
177 GCCGGGGCGG-----CCTGCTGCGCTGCTGAGACCATCTTCCGCG----- 222
340 TTGCTGAGAGTGTGCTGTGAGGACCATCTTCCGCG-----AAGAAACAACCCACCACCATCTG 397
223 TGC-----TGCTCTACGACAT--GAAC--GGCT-----GTTACTCA--CGCC-- 258
398 TGCATGACAAAGTGTCAACCTCATCCAGTCTGCGCTGACGCGTTCGCGAGTCCGCG 457
259 -TC---AAGGAGCTGTGCGCA-----CCCTGC-----CCC--AGAACCAG 293
458 ATCTGACAGTG--TGCTACCATCTATGAGGACCTTCGCGAGGAAGGCTGGAGTTCCT 516
294 AAGG-----TGAGCAAGGTGGAGATTTCTCCA-----GCAGTCTATGAGT--- 333
517 ATGACTGACCTGACATGTGTCAACCATCCACACACCCAGAGGACCGTGTTCAACTCAG 576
334 ----ACATCAGGAC---CTTCAGTTGGAGCTGAATCGGAATCCGAATTCGGACCCCG 386
577 AGACACATCAGGACAGGATTTCTGTTGGACATG--ACTC--CAGCCAGCAGAGGACTCT- 632
387 GGGGGCC--GAGGGCTCCGCTCCGCTCCGCTCAGCAG--CCTCAACGG-----CGAGATC 440
633 ---GGCCAGCATGCTGCCCTCTG---CCCGCCCGCCCATCTCTCTCGGTGACACGCG 686
441 AGCGCCTGACGCGCGGAGCGCATCGTTCCTGGG-----GACGATCGC----- 485
687 ATAGACCAAC--CCCGAACAGATTGGGAAGTGGCAGTGAGCTGGAGATGCTGAGTGG 745
486 -ATCTTG-----TGTC-----GTGA--AGC---GCTTCCCGCAG---GAGCGG 521
746 GAACTGTGAGGTGATGTCGAGATGTCGAGGAGTGTGTGTC--CACCAGGCGGAGCGCG 804
522 CGGACC-----CC-AGCCATCCAG-----GGGGCA 545
805 CAGACCTGGAGCTGCTGAGGAGCTCAACGCGCAGTGCAGGCGCATGCGAGCGGCTCC 864
546 AGAGGAATTACGTCTCTGTGGTC--TCCCGCAAGCGGCC--TC---GCGGATCTGAG 598
865 TGAGTGATACCTGTCTC--CGGGCCCATGCCCCAAGGAGCCCTTCAGAGCCACACTGCG 922
599 GGAGAACAGACCGATC---GGC--GGCCACTG-----CG-----CCC 631
923 AG---TCGAGGCTGCTGGAGGCTGGCCACAGTGGAAATTCGCGGACCTATTGTCTCC 979
632 TTAACTGTCATC--CAGCCTGGGG--CTGAGGCTGAGGCACTGGC----- 671
980 TACCCTGC-TCTGCTCATGCGGCGCCCATGTTTGG--CTGGCCACTGAGGTAGGTG 1036
672 --GAGGAGAGGCGCTCTCTGACACACTACTAGTC--ACCAGAGACTTTAGGGGTG 727
1037 TGGAGGTGTGAGGCGCC---CTGAGGAGCTGC--GGCGGCCCGAGGTAC---GAAGCTG 1087
728 GGATTCACCTCG-TGTG-----TTTCTATTTTTTGAAGAGCAGACATTTTAAATAATG 781
1088 CAACCTTGGCGCAGTGGGCGAGATCTCAT-----CAGCC-----CGAGCG 1128
782 CAGGTTTGTGCTTCTCAGATTTCTGAGG-----AAATTTGTTTGTATTGTA 828
1129 TGCAGGTGAGGCTTCAGGGGATGCTGGGCGCCCACTGCGCCCTCCGCTGCTTGCCTCCA 1188
```


QY 864 CAATAGTCTCTGGGGCTGT-TTTTTCGTATTAAACAATAATTAGATGCGAAAA 921
 Db 1249 CTCCTGTACTG-GGAGAGGTGCTTTTGT--ATCCCAATT--AAGGTAGAAAA 1298

RESULT 11
 ABU98259
 ID ABU98259 standard; protein; 1300 AA.
 XX AC ABU98259;
 XX DT 30-JUL-2003 (first entry)
 XX DE Novel human secreted and transmembrane protein #2.
 XX KW Human; secreted and transmembrane protein; PRO; cytotstatic; gene therapy;
 KW chondrocyte stimulator; tumour; adrenal tumour; lung tumour;
 KW colon tumour; breast tumour; prostate tumour; rectal tumour;
 KW Cervical tumour; liver tumour; chromosome identification.
 XX OS Homo sapiens.
 XX PN US2003017544-A1.
 XX PD 23-JAN-2003.
 XX PF 21-JUN-2002; 2002US-00176915.
 XX PR 18-SEP-1997; 97US-0059263P.
 PR 18-SEP-1997; 97US-0059266P.
 PR 17-OCT-1997; 97US-0062250P.
 PR 21-OCT-1997; 97US-0063486P.
 PR 24-OCT-1997; 97US-0063120P.
 PR 24-OCT-1997; 97US-0063121P.
 PR 28-OCT-1997; 97US-0063540P.
 PR 28-OCT-1997; 97US-0063541P.
 PR 28-OCT-1997; 97US-0063544P.
 PR 28-OCT-1997; 97US-0063564P.
 PR 29-OCT-1997; 97US-0063734P.
 PR 31-OCT-1997; 97US-0063870P.
 PR 31-OCT-1997; 97US-0064103P.
 PR 13-NOV-1997; 97US-0065311P.
 PR 21-NOV-1997; 97US-0066120P.
 PR 24-NOV-1997; 97US-0066466P.
 PR 24-NOV-1997; 97US-0066772P.
 PR 11-DEC-1997; 97US-0069335P.
 PR 12-DEC-1997; 97US-0069425P.
 PR 17-DEC-1997; 97US-0069870P.
 PR 18-DEC-1997; 97US-0068017P.
 PR 10-MAR-1998; 98US-0077450P.
 PR 11-MAR-1998; 98US-0077632P.
 PR 11-MAR-1998; 98US-0077649P.
 PR 20-MAR-1998; 98US-0078886P.
 PR 20-MAR-1998; 98US-0078939P.
 PR 27-MAR-1998; 98US-0079664P.
 PR 27-MAR-1998; 98US-0079786P.
 PR 31-MAR-1998; 98US-0080107P.
 PR 31-MAR-1998; 98US-0080194P.
 PR 01-APR-1998; 98US-0080327P.
 PR 01-APR-1998; 98US-0080333P.
 PR 08-APR-1998; 98US-0081049P.
 PR 08-APR-1998; 98US-0081070P.
 PR 09-APR-1998; 98US-0081195P.
 PR 15-APR-1998; 98US-0081838P.
 PR 21-APR-1998; 98US-0082568P.
 PR 21-APR-1998; 98US-0082569P.
 PR 22-APR-1998; 98US-0082704P.
 PR 22-APR-1998; 98US-0082797P.
 PR 28-APR-1998; 98US-0083322P.
 PR 29-APR-1998; 98US-0083495P.
 PR 29-APR-1998; 98US-0083496P.
 PR 29-APR-1998; 98US-0083499P.
 PR 29-APR-1998; 98US-0083559P.

PR 05-MAY-1998; 98US-0084366P.
 PR 06-MAY-1998; 98US-0084414P.
 PR 07-MAY-1998; 98US-0084639P.
 PR 07-MAY-1998; 98US-0084640P.
 PR 07-MAY-1998; 98US-0084643P.
 PR 15-MAY-1998; 98US-0085579P.
 PR 15-MAY-1998; 98US-0085580P.
 PR 15-MAY-1998; 98US-0085582P.
 PR 15-MAY-1998; 98US-0085700P.
 PR 18-MAY-1998; 98US-0086023P.
 PR 22-MAY-1998; 98US-0086392P.
 PR 22-MAY-1998; 98US-0086486P.
 PR 28-MAY-1998; 98US-0087098P.
 PR 28-MAY-1998; 98US-0087208P.
 PR 02-JUN-1998; 98US-0087609P.
 PR 02-JUN-1998; 98US-0087759P.
 PR 03-JUN-1998; 98US-0087827P.
 PR 04-JUN-1998; 98US-0088025P.
 PR 04-JUN-1998; 98US-0088028P.
 PR 04-JUN-1998; 98US-0088029P.
 PR 04-JUN-1998; 98US-0088033P.
 PR 04-JUN-1998; 98US-0088126P.
 PR 05-JUN-1998; 98US-0088167P.
 PR 05-JUN-1998; 98US-0088202P.
 PR 05-JUN-1998; 98US-0088212P.
 PR 05-JUN-1998; 98US-0088217P.
 PR 09-JUN-1998; 98US-0088555P.
 PR 10-JUN-1998; 98US-0088722P.
 PR 10-JUN-1998; 98US-0088738P.
 PR 10-JUN-1998; 98US-0088740P.
 PR 10-JUN-1998; 98US-0088811P.
 PR 10-JUN-1998; 98US-0088824P.
 PR 10-JUN-1998; 98US-0088825P.
 PR 10-JUN-1998; 98US-0088826P.
 PR 11-JUN-1998; 98US-0088861P.
 PR 11-JUN-1998; 98US-0088863P.
 PR 11-JUN-1998; 98US-0088876P.
 PR 12-JUN-1998; 98US-0089090P.
 PR 12-JUN-1998; 98US-0089105P.
 PR 16-JUN-1998; 98US-0089512P.
 PR 16-JUN-1998; 98US-0089514P.
 PR 17-JUN-1998; 98US-0089338P.
 PR 17-JUN-1998; 98US-0089398P.
 PR 17-JUN-1998; 98US-0089653P.
 PR 18-JUN-1998; 98US-0089908P.
 PR 19-JUN-1998; 98US-0089952P.
 PR 22-JUN-1998; 98US-0090246P.
 PR 22-JUN-1998; 98US-0090252P.
 PR 22-JUN-1998; 98US-0090254P.
 PR 24-JUN-1998; 98US-0090429P.
 PR 24-JUN-1998; 98US-0090435P.
 PR 24-JUN-1998; 98US-0090444P.
 PR 24-JUN-1998; 98US-0090461P.
 PR 24-JUN-1998; 98US-0090535P.
 PR 24-JUN-1998; 98US-0090540P.
 PR 25-JUN-1998; 98US-0090676P.
 PR 25-JUN-1998; 98US-0090678P.
 PR 25-JUN-1998; 98US-0090688P.
 PR 25-JUN-1998; 98US-0090690P.
 PR 25-JUN-1998; 98US-0090694P.
 PR 25-JUN-1998; 98US-0090695P.
 PR 25-JUN-1998; 98US-0090696P.
 PR 26-JUN-1998; 98US-00105413.
 PR 26-JUN-1998; 98US-0090862P.
 PR 26-JUN-1998; 98US-0090863P.
 PR 26-JUN-1998; 98US-0091010P.
 PR 01-JUL-1998; 98US-0091359P.
 PR 01-JUL-1998; 98US-0091544P.
 PR 02-JUL-1998; 98US-0091478P.
 PR 02-JUL-1998; 98US-0091486P.
 PR 02-JUL-1998; 98US-0091826P.
 PR 02-JUL-1998; 98US-0091828P.
 PR 02-JUL-1998; 98US-0091632P.

PR	04-AUG-1998;	98US-0095282P.	
PR	10-AUG-1998;	98US-0095998P.	
PR	10-AUG-1998;	98US-0096012P.	
PR	17-AUG-1998;	98US-0096757P.	
PR	17-AUG-1998;	98US-0096766P.	
PR	17-AUG-1998;	98US-0096867P.	
PR	17-AUG-1998;	98US-0096891P.	
PR	17-AUG-1998;	98US-0096897P.	
PR	18-AUG-1998;	98US-0096949P.	
PR	18-AUG-1998;	98US-0096959P.	
PR	18-AUG-1998;	98US-0097022P.	
PR	18-AUG-1998;	98US-0097952P.	
PR	26-AUG-1998;	98US-0097954P.	
PR	26-AUG-1998;	98US-0097955P.	
PR	26-AUG-1998;	98US-0097971P.	
PR	26-AUG-1998;	98US-0097974P.	
PR	01-SEP-1998;	98US-0098014P.	
PR	01-SEP-1998;	98US-0098716P.	
PR	01-SEP-1998;	98US-0098723P.	
PR	02-SEP-1998;	98US-0098803P.	
PR	02-SEP-1998;	98US-0098821P.	
PR	02-SEP-1998;	98US-0098843P.	
PR	09-SEP-1998;	98US-0099602P.	
PR	10-SEP-1998;	98US-0099741P.	
PR	10-SEP-1998;	98US-0099754P.	
PR	10-SEP-1998;	98US-0099763P.	
PR	10-SEP-1998;	98US-0099812P.	
PR	15-SEP-1998;	98US-0100388P.	
PR	16-SEP-1998;	98US-0100662P.	
PR	16-SEP-1998;	98US-0100664P.	
PR	16-SEP-1998;	98US-0101751P.	
PR	16-SEP-1998;	98US-01019330.	
PR	17-SEP-1998;	98US-0100683P.	
PR	17-SEP-1998;	98US-0100884P.	
PR	17-SEP-1998;	98US-0100919P.	
PR	17-SEP-1998;	98US-0100930P.	
PR	18-SEP-1998;	98US-0100849P.	
PR	18-SEP-1998;	98US-0101014P.	
PR	18-SEP-1998;	98US-0101068P.	
PR	23-SEP-1998;	98US-0101471P.	
PR	23-SEP-1998;	98US-0101472P.	
PR	23-SEP-1998;	98US-0101475P.	
PR	23-SEP-1998;	98US-0101477P.	
PR	24-SEP-1998;	98US-0101738P.	
PR	24-SEP-1998;	98US-0101739P.	
PR	24-SEP-1998;	98US-0101743P.	
PR	24-SEP-1998;	98US-0101922P.	
PR	25-SEP-1998;	98US-0101786P.	
PR	29-SEP-1998;	98US-0102207P.	
PR	29-SEP-1998;	98US-0102240P.	
PR	29-SEP-1998;	98US-0102330P.	
PR	29-SEP-1998;	98US-0102331P.	
PR	30-SEP-1998;	98US-0102487P.	
PR	30-SEP-1998;	98US-0102570P.	
PR	30-SEP-1998;	98US-0102571P.	
PR	01-OCT-1998;	98US-0102684P.	
PR	01-OCT-1998;	98US-0102687P.	

Query Match 43.2%; Score 2491; DB 6; Length 1300;

Best Local Similarity 42.6%; Pred. No. 3.5e-104;

Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

Qy	1	GGGGCCCATCTGTTTTCAGC-----CAGTCGCCAAG---AATCATGAAGTCGCCAGTG 51
Db	40	GTGGCAGCGGGTAGCAGCAATGGACTTTCTCTGGGGAACCCGTTTCAGCTCCAGTG 99
Qy	52	G--CAGCAC--CGCCACCGCGCGCGGC--CCAGCTGCGCGTGA--AGGCC----- 98
Db	100	GGACAGCGCATCGAGAAGCACAGATGGCTCCCTGCAGACGAGACTGGGCCCTCAAC 159
Qy	99	--GG-----CAAGACAG-----CGAGCG 114
Db	160	ATGAGATCTGCGACATCATCAACGACGAGGAGGTCCCAAGATGCCCTCCGAGCA 219

Qy	115	GT-----GCGGG-----CGAGGTGGTGC--CCTGT--CTGTC 142
Db	220	GTAAAGAGAGAAATCGTGGGAATAAGAACTTCCACGAGGTGATGCTGCTCTCACAGTC 279
Qy	143	TGAGCA-----GAGCGTGCCATCTCGCGT-----GCCGGGCG 176
Db	280	TTAGAAACCTGTGTCAAGAACTGCGGGCACCGCTCCACGTGCTGTGCGCCAGCCAGGAC 339
Qy	177	GCGGGGCGCG-----CCTGCTGCGCTGTGACGAGCAGCAGGTTAAACG 222
Db	340	TTCTGTGAGAGTGTGCTGTGAGGACCATCTCTGCCC--AAGAACAAACCCACCCACCATCG 397
Qy	223	TGC-----TGCTCTACGACAT--GAAC--GGCT-----GTTACTCA--CGCC-- 258
Db	398	TGCATGACAAAGTCTCAACCTCATCCAGTCTGCGGTGCGCGTTCGCGAGCTCGCCCG 457
Qy	259	-TC---AAGAGCTGTGTGCCA-----CCCTGC-----CCC--AGAACCGC 293
Db	458	ATCTGACAGGTG-TGGTCAACCATCTATGAGGACCTCGGAGGAAAGGCGCTGGAGTTCCCC 516
Qy	294	AAG-----TGAGCAGGTGGAGATTCTCCA-----GCACGTCACTGACT-- 333
Db	517	ATGACTGACCTGGACATGCTGTCAACCATCCACACCCAGAGGACGGTGTCAACTCAG 576
Qy	334	---ACATCAGGGAC---CTTCAGTTGGAGCTGAATCGGAATCCGAAGTTGGGACCCCC 386
Db	577	AGACACAATCAGACAGGATTCTGTGGCACTG-ACTC--CAGCCACGACAGGACTCT- 632
Qy	387	GGGGGCC-GAGGGCTGCGGTGCGGCTCGCTCAGCAG--CCTCAACG--CGAGATC 440
Db	633	---GGCCAGCATGTCCTCTCTG---CCGCCCCGCCCATCTCTCCCGTGAACGCC 686
Qy	441	AGCGCCTGACGGCCGAGGGCGGATGCTTCCTGCG-----GACGATGCG----- 485
Db	687	ATAGCACCAAC-CCCGAAACAGATTGGGAAGTGGCAGTGAGTGGAGATGCTGAGTGG 745
Qy	486	-ATCTTG-----TGTC-----GCTGA--AGC-----GCCTCCCCCAG--GGACCG 521
Db	746	GAACGTGAGGGTGATGTCCGAGATGCTGACGAGTGTGTGCC--CACCAGGCCAGCCCG 804
Qy	522	CGGACC-----CC-AGCATCCAG-----GGGGCA 545
Db	805	CAGACCTGGAGCTGCTGCAGGAGCTCAACCGCAGTCCGAGCCATCAGCAGCGGGTCC 864
Qy	546	AGAGGAATTAGTCTCTGTGGGTGTC--TCCCCAACCGCC--TC---GGCGGATCTGAG 598
Db	865	TGAGTGATACCTCTCTC--CGGGCCCATGCCCCAAGAGGCCCTTCAGAGCCACACTGCC 922
Qy	599	GGAGAACAAAGACCCGATC---GGC-GGCCACTG-----CG-----CCC 631
Db	923	AG---TCGAGGCTTGGCTGGAGGCTGGCCACAGTGGAAATTCGCCGAGCCTATTGTCCC 979
Qy	632	TTAACTGCATC--CAGCTGGGG--CTGAGCTGAGGCACTGGC----- 671
Db	980	TACCTGC-TCTGCTGATGGGCCCCCATGGCTTTGG--CTGGCCACTGAGGGTAGGGTG 1036
Qy	672	--GAGGAGGGGCTCTCTCTCAGACCTACTAGTGC--ACAGAGACTTTAGGGGTG 727
Db	1037	TGGAGGTGTGAGGCCCC--CTGAGGAGTGC--GGGGCCCAAGTAC-----GAAGTGT 1087
Qy	728	GGATTCCACTCG-TGTG-----TTTCTATTTTGAAGCAGACATTTTAAATAATGTT 781
Db	1088	CAACTCTCGCGCAGTGGCGGAGATCTCAT-----CAGCC-----CCAGGC 1128
Qy	782	CACGTTTGGTCTCTCAGATTTCTGAG-----AAATTGCTTTGTATTGTA 828
Db	1129	TGAGGTGAGGCTTCAGGGGATGCTGGGGCCCCACTGCCCTCGCTGCTTGCCTCCCA 1188
Qy	829	T-ATTACAAATGATC-----ACCG-----ACTGAG-----AATATGTTT--TA 863
Db	1189	TCCTTCTCTGTTCTTCTGCGCGGGCACCACAGCACTGGGGCTCACCTCTTGTGATC 1248

RESULT 10	PR	06-MAY-1998;	98US-0084414P.
ABU99044	PR	07-MAY-1998;	98US-0084639P.
ID ABU99044 standard; protein; 1300 AA.	PR	07-MAY-1998;	98US-0084640P.
XX	PR	07-MAY-1998;	98US-0084643P.
XX	PR	15-MAY-1998;	98US-008579P.
XX	PR	15-MAY-1998;	98US-0085580P.
XX	PR	15-MAY-1998;	98US-0085582P.
XX	PR	15-MAY-1998;	98US-0085700P.
XX	PR	18-MAY-1998;	98US-0086023P.
XX	PR	22-MAY-1998;	98US-0086392P.
XX	PR	22-MAY-1998;	98US-0086486P.
XX	PR	28-MAY-1998;	98US-0087098P.
XX	PR	28-MAY-1998;	98US-0087208P.
XX	PR	02-JUN-1998;	98US-0087609P.
XX	PR	02-JUN-1998;	98US-0087759P.
XX	PR	03-JUN-1998;	98US-0088025P.
XX	PR	04-JUN-1998;	98US-0088028P.
XX	PR	04-JUN-1998;	98US-0088029P.
XX	PR	04-JUN-1998;	98US-0088033P.
XX	PR	04-JUN-1998;	98US-0088326P.
XX	PR	05-JUN-1998;	98US-0088167P.
XX	PR	05-JUN-1998;	98US-0088202P.
XX	PR	05-JUN-1998;	98US-0088212P.
XX	PR	05-JUN-1998;	98US-0088217P.
XX	PR	09-JUN-1998;	98US-0088655P.
XX	PR	10-JUN-1998;	98US-0088722P.
XX	PR	10-JUN-1998;	98US-0088738P.
XX	PR	10-JUN-1998;	98US-0088740P.
XX	PR	10-JUN-1998;	98US-0088811P.
XX	PR	10-JUN-1998;	98US-0088824P.
XX	PR	10-JUN-1998;	98US-0088825P.
XX	PR	10-JUN-1998;	98US-0088826P.
XX	PR	11-JUN-1998;	98US-0088861P.
XX	PR	11-JUN-1998;	98US-0088863P.
XX	PR	11-JUN-1998;	98US-0088876P.
XX	PR	12-JUN-1998;	98US-0089090P.
XX	PR	12-JUN-1998;	98US-0089105P.
XX	PR	16-JUN-1998;	98US-0089512P.
XX	PR	16-JUN-1998;	98US-0089514P.
XX	PR	17-JUN-1998;	98US-0089538P.
XX	PR	17-JUN-1998;	98US-0089598P.
XX	PR	17-JUN-1998;	98US-0089653P.
XX	PR	18-JUN-1998;	98US-0089908P.
XX	PR	19-JUN-1998;	98US-0089952P.
XX	PR	22-JUN-1998;	98US-0090246P.
XX	PR	22-JUN-1998;	98US-0090252P.
XX	PR	22-JUN-1998;	98US-0090254P.
XX	PR	24-JUN-1998;	98US-0090429P.
XX	PR	24-JUN-1998;	98US-0090435P.
XX	PR	24-JUN-1998;	98US-0090444P.
XX	PR	24-JUN-1998;	98US-0090461P.
XX	PR	24-JUN-1998;	98US-0090535P.
XX	PR	24-JUN-1998;	98US-0090540P.
XX	PR	25-JUN-1998;	98US-0090676P.
XX	PR	25-JUN-1998;	98US-0090678P.
XX	PR	25-JUN-1998;	98US-0090688P.
XX	PR	25-JUN-1998;	98US-0090690P.
XX	PR	25-JUN-1998;	98US-0090694P.
XX	PR	25-JUN-1998;	98US-0090695P.
XX	PR	25-JUN-1998;	98US-0090696P.
XX	PR	26-JUN-1998;	98US-0010543.
XX	PR	26-JUN-1998;	98US-0090862P.
XX	PR	26-JUN-1998;	98US-0090863P.
XX	PR	26-JUN-1998;	98US-0091010P.
XX	PR	01-JUL-1998;	98US-0091359P.
XX	PR	01-JUL-1998;	98US-0091544P.
XX	PR	02-JUL-1998;	98US-0091478P.
XX	PR	02-JUL-1998;	98US-0091486P.
XX	PR	02-JUL-1998;	98US-0091656P.
XX	PR	02-JUL-1998;	98US-0091628P.
XX	PR	02-JUL-1998;	98US-0091632P.
XX	PR	24-JUL-1998;	98US-0094006P.

ABU99044	Human; secreted and transmembrane protein #2.
XX	Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy;
XX	chondrocyte stimulator; tumour; adrenal tumour; lung tumour;
XX	colon tumour; breast tumour; prostate tumour; rectal tumour;
XX	cervical tumour; liver tumour; TNF-alpha release;
XX	tumour necrosis factor alpha release; chondrocyte cell proliferation;
XX	chondrocyte cell differentiation; pharmaceutical; diagnostic; biosensor;
XX	bioindicator.
OS	Homo sapiens.
XX	US2003013153-A1.
XX	16-JAN-2003.
XX	19-JUN-2002; 2002US-00175737.
XX	18-SEP-1997; 97US-0059263P.
XX	18-SEP-1997; 97US-0059266P.
XX	17-OCT-1997; 97US-0062250P.
XX	21-OCT-1997; 97US-0063486P.
XX	24-OCT-1997; 97US-0063120P.
XX	24-OCT-1997; 97US-0063540P.
XX	28-OCT-1997; 97US-0063541P.
XX	28-OCT-1997; 97US-0063544P.
XX	28-OCT-1997; 97US-0063564P.
XX	29-OCT-1997; 97US-0063734P.
XX	31-OCT-1997; 97US-0063870P.
XX	31-OCT-1997; 97US-0064103P.
XX	13-NOV-1997; 97US-0065311P.
XX	21-NOV-1997; 97US-0066120P.
XX	24-NOV-1997; 97US-0066466P.
XX	24-NOV-1997; 97US-0066772P.
XX	11-DEC-1997; 97US-0069335P.
XX	12-DEC-1997; 97US-0069425P.
XX	17-DEC-1997; 97US-0069870P.
XX	18-DEC-1997; 97US-0068017P.
XX	10-MAR-1998; 98US-0077450P.
XX	11-MAR-1998; 98US-0077632P.
XX	20-MAR-1998; 98US-0078986P.
XX	20-MAR-1998; 98US-0078939P.
XX	27-MAR-1998; 98US-0079664P.
XX	27-MAR-1998; 98US-0079786P.
XX	31-MAR-1998; 98US-0080107P.
XX	31-MAR-1998; 98US-0080194P.
XX	01-APR-1998; 98US-0080327P.
XX	01-APR-1998; 98US-0080333P.
XX	08-APR-1998; 98US-0081049P.
XX	08-APR-1998; 98US-0081070P.
XX	09-APR-1998; 98US-0081195P.
XX	15-APR-1998; 98US-0081838P.
XX	21-APR-1998; 98US-0082568P.
XX	21-APR-1998; 98US-0082569P.
XX	22-APR-1998; 98US-0082704P.
XX	22-APR-1998; 98US-0082797P.
XX	28-APR-1998; 98US-0083322P.
XX	29-APR-1998; 98US-0083495P.
XX	29-APR-1998; 98US-0083496P.
XX	29-APR-1998; 98US-0083499P.
XX	05-MAY-1998; 98US-0083559P.
XX	05-MAY-1998; 98US-0084366P.

17-AUG-1998; 98US-0096897P.
 18-AUG-1998; 98US-0096949P.
 18-AUG-1998; 98US-0096959P.
 18-AUG-1998; 98US-0097022P.
 26-AUG-1998; 98US-0097952P.
 26-AUG-1998; 98US-0097954P.
 26-AUG-1998; 98US-0097955P.
 26-AUG-1998; 98US-0097971P.
 26-AUG-1998; 98US-0097974P.
 26-AUG-1998; 98US-0098014P.
 01-SEP-1998; 98US-0098716P.
 01-SEP-1998; 98US-0098723P.
 01-SEP-1998; 98US-0098803P.
 02-SEP-1998; 98US-0098821P.
 02-SEP-1998; 98US-0098843P.
 09-SEP-1998; 98US-0099602P.
 10-SEP-1998; 98US-0099741P.
 10-SEP-1998; 98US-0099754P.
 10-SEP-1998; 98US-0099812P.
 15-SEP-1998; 98US-0100388P.
 16-SEP-1998; 98US-0100682P.
 16-SEP-1998; 98US-0100684P.
 16-SEP-1998; 98US-0101751P.
 16-SEP-1998; 98US-0101751P.
 17-SEP-1998; 98US-0100683P.
 17-SEP-1998; 98US-0100684P.
 17-SEP-1998; 98US-0100919P.
 17-SEP-1998; 98US-0100930P.
 18-SEP-1998; 98US-0100849P.
 18-SEP-1998; 98US-0101044P.
 18-SEP-1998; 98US-0101068P.
 23-SEP-1998; 98US-0101471P.
 23-SEP-1998; 98US-0101472P.
 23-SEP-1998; 98US-0101475P.
 23-SEP-1998; 98US-0101477P.
 24-SEP-1998; 98US-0101738P.
 24-SEP-1998; 98US-0101739P.
 24-SEP-1998; 98US-0101743P.
 24-SEP-1998; 98US-0101922P.
 25-SEP-1998; 98US-0101786P.
 29-SEP-1998; 98US-0102207P.
 29-SEP-1998; 98US-0102240P.
 29-SEP-1998; 98US-0102330P.
 29-SEP-1998; 98US-0102331P.
 30-SEP-1998; 98US-0102487P.
 30-SEP-1998; 98US-0102570P.
 30-SEP-1998; 98US-0102571P.
 01-OCT-1998; 98US-0102684P.
 01-OCT-1998; 98US-0102687P.
 02-OCT-1998; 98US-0102965P.
 06-OCT-1998; 98US-0103258P.
 06-OCT-1998; 98US-0103449P.

Query Match 43.2%; Score 2491; DB 6; Length 1300;
 Best Local Similarity 42.6%; Pred. No. 3.5e-104; Indels 458; Gaps 83;
 Matches 562; Conservative 0; Mismatches 299;
 1 GGGGCCCATCTGTTTCAGC-----CAGTGGCAAG---AATCATGAAGTCCGAGTG 51
 40 GTGGCAGCGCGGTGACGACAAATGAACTTTCTCTGGGGAACCCGTTTCACTCTCCAGTG 99
 52 G--CAGCAC--CGCCACCGCCCGCGGC--CCAGCTGGCGCTGA--AGGCC----- 98
 100 GACAGCGCATCGAGAAGCCACAGATGGCTCCCTGACAGACGAGACTGGGCCCTCAAC 159
 99 --GG-----CAAGACAG-----CGAGCG 114
 160 ATGGAGATCTGGACATCATCAAGACGAGAGGAGAGGTCCTCCAAAGATGCCCTCCGAGCA 219
 115 GT-----GCGGG-----CGAGTGGTGC--GCTGT--CTGTC 142
 220 GTAAGAGAGAGATCTGGGGAATAGAACTTCCACGAGGTGATGCTGGCTCTCAGATC 279

143 TGAGCA-----GAGCTGGCCATCTCGCGCT-----GCGGGGC 176
 280 TTAGAAACCTGTGTCAAGAACTGCGGCACCGCTTCCACGCTGCTGGTGGCCAGCAGAC 339
 177 GCGGGGCGG-----CCTGCTGCGCTGCTGGACGAGCAGCAGGTAACG 222
 340 TTGCTGGAGAGTGTGCTGTGAGGACCACTCTGCGCC--AAGAACAACCCACCACTCG 397
 223 TGC-----TGCTCTACGACAT---GAAC--GGCT-----GTTACTCA--CGCC-- 258
 398 TGCATGACAAAGTGTCAACCTCATCCAGTCTGGGCTGACGCGTTCGCGAGCTGCGCG 457
 259 -TC---AAGGAGCTGGTCCCA-----CCCTGC-----GCC--AGNACGC 293
 458 ATCTGACAGTG--TGCTACCATCTATGAGGACCTCGGAGAAAGCGCTGAGTTCGCC 516
 294 AAGG-----TGAGCAAGTGGAGATTCTCCA-----GACGTCTACGACT--- 333
 517 ATGACTGACCTGGACATGCTGTACCCATCCACACCCAGAGGACCGTGTCACTCAG 576
 334 ----ACATCAGGAC---CTTCAGTTGAGCTGAACTCGGATCCGAAGTTGGGACCC 386
 577 AGACAAATCAGGACAGGATTCTGTTGGCACTG--ACTC--CAGCCAGCAAGAGGACTCT-- 632
 633 ---GGCCAGCATGCTGCCCTCTG---CCGCGCCCGCCCATACTCTCGGTGACACGCC 686
 441 AGCGCCTGAGCGCGGAGCGGCATGCGTTCCTGGG-----GAGCATGC----- 485
 687 ATAGCAGCAAC--CCCGAACAATGAGGAGCTGGCAGTGGAGTGAGTGGTGGTGG 745
 486 -ATCTTG-----TGTC-----GCTGA--AGC---GCTTCCCCCAG--GGACGG 521
 746 GAACTGAGGGTGTGTCGAGATGCTGACGAGTGTGGTGC--CACCAGGCGGAGCGCG 804
 522 CGGACC-----CC-AGCCATCAG---GGGSCA 545
 805 CAGACTGGAGCTGTGAGGAGCTCAACCGCACGTCGCGAGCTATGACGAGCGGCTCC 864
 546 AGAGGAATACGTGCTCTGTGGTC--TCCGCCAACGCGCC--TC---GCGGATCTGAG 598
 865 TGAGTGATACCTGTCTC--CGGCGCCATGCCCCAAGAGAGCCCTTCAGAGCCACACTGCC 922
 599 GGAGAACAGACCGATC---GSC--GGCCACTG-----CG-----CCC 631
 923 AG--TCGAGGCTGCTGGAGGCTGGTCACTGGAATCTGCGGAGCCTATTGTCCC 979
 632 TTAATCTGATC--CAGCTTGGG--CTGAGGTGAGGCACTGGC----- 671
 980 TACCCTGC-TCTGCTGATGGGCGCCCATGGCTTTGG--CTGGCCACTGAGGTTAGG 1036
 672 --GAGGAGAGGGGCTCTCTGACACACTACTAGTC--ACCAGAGACTTTAGGGGCTG 727
 1037 TGAGGTGTGAGGCCCC--CTGAGGAGCTGC--GGCGGCCAGGTAC---GAAGCTG 1087
 728 GGAATTCACCTCG-TGTG-----TTTCTATTTTGAAGAGACATTTTAAATAATG 781
 1088 CAACTCTGCGCGAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC 1128
 782 CAGTTTGGTGTCTCTCAGATTCTCAGG-----AAATGCTTTGTTATTGTA 828
 1129 TGCAGGTGAGCTTCAGGGGATGCTGGGCCCCCACTGCCCTCCGCTGCCCTGCCCTCA 1188
 829 T-ATTACATGATC-----ACCG-----ACTGAG-----AATATTGTTT--TA 863
 1189 TCCTCTCTGTTCTCTTCTGGCGGGGACCAAGAGCTGGGGCTCACCTCTTGGTTGATC 1248
 864 CAATAGTCTTGGGGCTGT-TTTTTTGTATTAACAAATAATTTAGATGTTGTAATAA 921
 1249 CTCTTGTACTG-GGAGAGGTGCTTTTGT--ATCCCCAATTA-----AAGGTAGAAA 1298

XX AC ABU96435;
XX DT 25-JUL-2003 (first entry)
XX DE Novel human secreted and transmembrane protein #2.
XX KW Human; secreted and transmembrane protein; PRO; transgenic animal;
KW knockout; chromosome identification; tissue typing; tumour;
KW chondrocyte proliferation; chondrocyte differentiation;
KW tumor necrosis factor-alpha release stimulator.
XX OS Homo sapiens.
XX PN US2003036144-A1.
XX PD 20-FEB-2003.
XX PF 01-JUL-2002; 2002US-00187601.
XX PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063545P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066456P.
PR 24-NOV-1997; 97US-0066722P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 11-MAR-1998; 98US-0077886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081135P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085586P.
PR 15-MAY-1998; 98US-0085591P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 28-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088266P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.

PR	26-AUG-1998;	98US-0097954P.
PR	26-AUG-1998;	98US-0097955P.
PR	26-AUG-1998;	98US-0097971P.
PR	26-AUG-1998;	98US-0097974P.
PR	26-AUG-1998;	98US-0098014P.
PR	01-SEP-1998;	98US-0098716P.
PR	01-SEP-1998;	98US-0098723P.
PR	02-SEP-1998;	98US-0098803P.
PR	02-SEP-1998;	98US-0098821P.
PR	02-SEP-1998;	98US-0098843P.
PR	09-SEP-1998;	98US-0099602P.
PR	10-SEP-1998;	98US-0099741P.
PR	10-SEP-1998;	98US-0099754P.
PR	10-SEP-1998;	98US-0099763P.
PR	10-SEP-1998;	98US-0099812P.
PR	15-SEP-1998;	98US-0100389P.
PR	16-SEP-1998;	98US-0100662P.
PR	16-SEP-1998;	98US-0100664P.
PR	16-SEP-1998;	98US-0101751P.
PR	16-SEP-1998;	98WO-US019330.
PR	17-SEP-1998;	98US-0100683P.
PR	17-SEP-1998;	98US-0100684P.
PR	17-SEP-1998;	98US-0100919P.
PR	17-SEP-1998;	98US-0100930P.
PR	18-SEP-1998;	98US-0100849P.
PR	18-SEP-1998;	98US-0101014P.
PR	18-SEP-1998;	98US-0101068P.
PR	23-SEP-1998;	98US-0101471P.
PR	23-SEP-1998;	98US-0101472P.
PR	23-SEP-1998;	98US-0104752P.
PR	23-SEP-1998;	98US-0104777P.
PR	24-SEP-1998;	98US-0101738P.
PR	24-SEP-1998;	98US-0101739P.
PR	24-SEP-1998;	98US-0101743P.
PR	24-SEP-1998;	98US-0101922P.
PR	25-SEP-1998;	98US-0101786P.
PR	29-SEP-1998;	98US-0102207P.
PR	29-SEP-1998;	98US-0102240P.
PR	29-SEP-1998;	98US-0102330P.
PR	29-SEP-1998;	98US-0102331P.
PR	30-SEP-1998;	98US-0102487P.
PR	30-SEP-1998;	98US-0102570P.
PR	30-SEP-1998;	98US-0102571P.
PR	01-OCT-1998;	98US-0102684P.
PR	01-OCT-1998;	98US-0102687P.
PR	02-OCT-1998;	98US-0102965P.
PR	06-OCT-1998;	98US-0103258P.
PR	06-OCT-1998;	98US-0103449P.
PR	07-OCT-1998;	98US-0016897B.

Query Match 43.2%; Score 2491; DB 6; Length 1300;

Best Local Similarity 42.6%; Pred. No. 3.5e-104;

Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

QY	1	GGGGCCCATCTGTTTCAGC-----CAGTCGCCAAG---AATCATGAAAGTCGCCAGTG 51
DB	40	GTGGCAGCGCGGTAGCAGCAATGGACTTTCTCTGGGAAACCGTTTCAGCTCTCCAGTG 99
QY	52	G--CAGCAC--CGCCACCGCGCGCGGC--CCAGCTGCGCGCTGA--AGGCC----- 98
DB	100	GGACAGCGCATCGAGAAAGCCACAGATGGCTCCCTGCGAGCGAGGACTGGGCGCTCAAC 159
QY	99	--CG-----CAAGACAG-----CGAGCG 114
DB	160	ATGGAGATCGGACATCATACAGACGGAGGAGGTCCCAAGATGCCCTCCGAGCA 219
QY	115	GT-----CGCGG-----CGAGTGGTGG--GCTGT--CTGTC 142
DB	220	GTAAGAGAGAGATCGTGGGATTAAGAACTTCCACAGGTGATGTGCTCTCAGTC 279
QY	143	TGAGCA-----GAGCGTGGCCATCTCGCGT-----GCCGGGCG 176
DB	280	TTAGAAACCTGTGTCAAGAACTCGGGCACCCTTCCACGTGTGTGTGGCCAGCAGAC 339

QY	177	GCGGGGCGCG-----CCTGCGTCCCTGCTGGACGAGCAGGAGTAAACG 222
DB	340	TTCTGTGGAGAGTGTCTGTGTGAGGACCATCTCGCCC--AAGAACAAACCACCCACCATCG 397
QY	223	TGC-----TGCTTACGACAT---GAAC--GGCT-----GTTACTCA---CGCC-- 258
DB	398	TGCATGACAAAGTGTCTACCTCATCCAGTCTGGGCTGACGCGTTCCGACGTGCGCG 457
QY	259	-TC---AAGGAGCTGGTGCCCA-----CCCTGC-----CCC--AGAACCGC 293
DB	458	ATCTGACAGTG-TGTTACCATCTATAGGAGACCTGCGGAGGAAAGCGCTGGAGTTCGCC 516
QY	294	AAAG-----TGAGCAAGGTGGAGATTCTCCA-----GCACGTCTACGACT--- 333
DB	517	ATGACTGACCTGGACATGTGTCAACCATCCACACCCAGAGGACCGTGTTCAACTCAG 576
QY	334	----ACATCAGGAC---CTTCAGTTGAGCTGAACTCGGAATCGAAGTTGGGACCCCG 386
DB	577	AGACACAATCAGGACAGGAGTTCTGTGGCACTG-ACTC--CAGCCAGCAAGAGGACTCT- 632
QY	387	GGGGGCC-GAGGGCTCGCGTCCGGCTCCGCTCAGCAC--CCTCAAGCG---CGAGATC 440
DB	633	---GGCCAGCATGCTGCCCTCTG---CCGCCCCGCCCATACTCTCCGGTGACACGCC 686
QY	441	AGCGCCCTGACGGCGAGCGCGGATGCTTCCTGGG-----GACATCGC----- 485
DB	687	ATAGCACCAAC-CCCGAAACAGATTGGGAAGCTGCGCAGTGAGCTGGAGATGTTGAGTGG 745
QY	486	-ATCTTG-----TGTC-----GCTGA---AGC---GCCTCCCCCAG---GGACCGG 521
DB	746	GAACGTGAGGGTGATGTCGAGATGCTGACGAGGTGTGTGCC-CACCCAGGCGGAGCCCG 804
QY	522	CGGACC-----CC-AGCCATCCAG---GGGGCA 545
DB	805	CAGACCTGGAGCTGTGTCAGGAGCTCAACCGCAGCTGCCGAGCCATGTCAGCAGCGGTCC 864
QY	546	AGAGGAATTACGTCTCTGTGGTG--TCCCCCAACGCGCC--TC---GCCGATCTGAG 598
DB	865	TGAGTGATACCTGCTC---CGGCCCCATGCCCCAGAGAGCCCTTCAGAGCCACACTGCC 922
QY	599	GGAGAAACAAGACCGATC---GGC-GGCCACTG-----CG-----CCC 631
DB	923	AG---TCGAGGCTGCTGGAGGCTGCCACAGTGGAAATTTCTGCGAGACTATTGTCCC 979
QY	632	TTAACTGCTATC--CAGCCTGGGG--CTGAGGCTGAGGCACTGGC----- 671
DB	980	TACCCCTGC-TCTGCTGATGGGCCCCCATGGCTTTGG--CTGGCCACTGAGGTAGGGTG 1036
QY	672	--GAGGAGAGGGCGCTCTCTGACACACTACTAGTC--ACCAGAGACTTTAGGGGGTG 727
DB	1037	TGAGGTTGGAGGCCCC---CTGAGGAGTGC--GGCGGCCAGGTAC---GAAGCTG 1087
QY	728	GGATTCACTCG-TGTG-----TTTCTATTTTTGAAGAGCAGACATTTTAAATAATGTT 781
DB	1088	CAACTCTGCGCGAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC 1128
QY	782	CAGTTTGGTCTTCTCAGATTTCTGAGG-----AAATTGCTTTGTATTGTA 828
DB	1129	TGAGGTGAGCTTCAGGGATGCTGGGGCCCCACTGCCCTCGCTGCTGCTGCCCTCCA 1188
QY	829	T-ATTACATGATC-----ACCG-----ACTGAG-----AATATTGTTT--TA 863
DB	1189	TCCTTCCTCTGTTCTTCTGCGCGGCAACACAGCACTGGGCTCACCTCTTGGTGATC 1248
QY	864	CAATAGTCTGTGGGCTGT-TTTTGTGTTATTAACAATAATTTAGATGGTGAAAAA 921
DB	1249	CTCTGTACTG-GGAGAGGTGCTTTTGT--ATCCCCAATTA-----AAGGTAGAAAA 1298

RESULT 9
ABU96435
ID ABU96435 standard; protein; 1300 AA.

XX DE Novel human secreted and transmembrane PRO protein #2.
XX DE Human; gene therapy; tissue typing; tumour; chondrocyte proliferation;
XX KW chondrocyte differentiation; tumour necrosis factor-alpha release;
XX KW affinity purification.
XX OS Homo sapiens.
XX PN US2003036147-A1.
XX PD 20-FEB-2003.
XX PF 02-JUL-2002; 2002US-00187741.
XX PR 18-SEP-1997; 97US-0059263P.
XX PR 18-SEP-1997; 97US-0059266P.
XX PR 17-OCT-1997; 97US-0062250P.
XX PR 21-OCT-1997; 97US-0063486P.
XX PR 24-OCT-1997; 97US-0063120P.
XX PR 24-OCT-1997; 97US-0063121P.
XX PR 28-OCT-1997; 97US-0063540P.
XX PR 28-OCT-1997; 97US-0063541P.
XX PR 28-OCT-1997; 97US-0063544P.
XX PR 28-OCT-1997; 97US-0063564P.
XX PR 29-OCT-1997; 97US-0063734P.
XX PR 31-OCT-1997; 97US-0063870P.
XX PR 31-OCT-1997; 97US-0064103P.
XX PR 13-NOV-1997; 97US-0065311P.
XX PR 21-NOV-1997; 97US-0066120P.
XX PR 24-NOV-1997; 97US-0066466P.
XX PR 24-NOV-1997; 97US-0066772P.
XX PR 11-DEC-1997; 97US-0069335P.
XX PR 12-DEC-1997; 97US-0069425P.
XX PR 17-DEC-1997; 97US-0069870P.
XX PR 18-DEC-1997; 97US-0068017P.
XX PR 10-MAR-1998; 98US-0077450P.
XX PR 11-MAR-1998; 98US-0077532P.
XX PR 11-MAR-1998; 98US-0077549P.
XX PR 20-MAR-1998; 98US-0078886P.
XX PR 20-MAR-1998; 98US-0078939P.
XX PR 27-MAR-1998; 98US-0079664P.
XX PR 27-MAR-1998; 98US-0079786P.
XX PR 31-MAR-1998; 98US-0080107P.
XX PR 31-MAR-1998; 98US-0080194P.
XX PR 01-APR-1998; 98US-0080327P.
XX PR 01-APR-1998; 98US-0080333P.
XX PR 08-APR-1998; 98US-0081049P.
XX PR 08-APR-1998; 98US-0081070P.
XX PR 09-APR-1998; 98US-0081195P.
XX PR 15-APR-1998; 98US-0081838P.
XX PR 21-APR-1998; 98US-0082568P.
XX PR 21-APR-1998; 98US-0082569P.
XX PR 22-APR-1998; 98US-0082704P.
XX PR 22-APR-1998; 98US-0082797P.
XX PR 28-APR-1998; 98US-0083222P.
XX PR 29-APR-1998; 98US-0083495P.
XX PR 29-APR-1998; 98US-0083496P.
XX PR 29-APR-1998; 98US-0083499P.
XX PR 29-APR-1998; 98US-0083559P.
XX PR 05-MAY-1998; 98US-0084366P.
XX PR 06-MAY-1998; 98US-0084414P.
XX PR 07-MAY-1998; 98US-0084639P.
XX PR 07-MAY-1998; 98US-0084640P.
XX PR 07-MAY-1998; 98US-0084643P.
XX PR 15-MAY-1998; 98US-0085579P.
XX PR 15-MAY-1998; 98US-0085580P.
XX PR 15-MAY-1998; 98US-0085582P.
XX PR 15-MAY-1998; 98US-0085700P.
XX PR 18-MAY-1998; 98US-0086023P.
XX PR 22-MAY-1998; 98US-0086392P.
XX PR 22-MAY-1998; 98US-0086486P.
XX PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 05-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 11-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 15-JUN-1998; 98US-0089512P.
PR 15-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090425P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 24-JUL-1998; 98US-0091632P.
PR 04-AUG-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095988P.
PR 17-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 18-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096945P.
PR 18-AUG-1998; 98US-0096959P.
PR 26-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.

KW tumour necrosis factor-alpha release; TNF-alpha release;
 KW chondrocyte proliferation; chondrocyte differentiation; tumour;
 KW adrenal tumour; lung tumour; colon tumour; breast tumour;
 KW prostate tumour; rectal tumour; cervical tumour; liver tumour.

XX Homo sapiens.

XX US2003032127-A1.

XX PD 13-FEB-2003.

XX PF 26-JUN-2002; 2002US-00183012.

XX PR 18-SEP-1997; 97US-0059263P.

XX PR 18-SEP-1997; 97US-0059266P.

XX PR 17-OCT-1997; 97US-0062250P.

XX PR 21-OCT-1997; 97US-0063486P.

XX PR 24-OCT-1997; 97US-0063120P.

XX PR 24-OCT-1997; 97US-0063121P.

XX PR 28-OCT-1997; 97US-0063540P.

XX PR 28-OCT-1997; 97US-0063541P.

XX PR 28-OCT-1997; 97US-0063544P.

XX PR 28-OCT-1997; 97US-0063564P.

XX PR 29-OCT-1997; 97US-0063734P.

XX PR 31-OCT-1997; 97US-0064103P.

XX PR 31-OCT-1997; 97US-0064103P.

XX PR 13-NOV-1997; 97US-0065311P.

XX PR 21-NOV-1997; 97US-0066120P.

XX PR 24-NOV-1997; 97US-0066466P.

XX PR 24-NOV-1997; 97US-0066772P.

XX PR 11-DEC-1997; 97US-0069335P.

XX PR 12-DEC-1997; 97US-0069425P.

XX PR 17-DEC-1997; 97US-0069870P.

XX PR 18-DEC-1997; 97US-0068017P.

XX PR 10-MAR-1998; 98US-0077450P.

XX PR 11-MAR-1998; 98US-0077632P.

XX PR 11-MAR-1998; 98US-0077649P.

XX PR 20-MAR-1998; 98US-0078886P.

XX PR 20-MAR-1998; 98US-0078939P.

XX PR 27-MAR-1998; 98US-0079564P.

XX PR 27-MAR-1998; 98US-0079786P.

XX PR 31-MAR-1998; 98US-0080107P.

XX PR 31-MAR-1998; 98US-0080194P.

XX PR 01-APR-1998; 98US-0080327P.

XX PR 01-APR-1998; 98US-0080333P.

XX PR 08-APR-1998; 98US-0081049P.

XX PR 08-APR-1998; 98US-0081070P.

XX PR 09-APR-1998; 98US-0081195P.

PR 02-JUN-1998; 98US-0087759P.
 PR 03-JUN-1998; 98US-0087827P.
 PR 04-JUN-1998; 98US-0088025P.
 PR 04-JUN-1998; 98US-0088028P.
 PR 04-JUN-1998; 98US-0088029P.
 PR 04-JUN-1998; 98US-0088033P.
 PR 04-JUN-1998; 98US-0088326P.
 PR 05-JUN-1998; 98US-0088167P.
 PR 05-JUN-1998; 98US-0088202P.
 PR 05-JUN-1998; 98US-0088212P.
 PR 05-JUN-1998; 98US-0088217P.
 PR 05-JUN-1998; 98US-0088655P.
 PR 09-JUN-1998; 98US-0088722P.
 PR 10-JUN-1998; 98US-0088738P.
 PR 10-JUN-1998; 98US-0088740P.
 PR 10-JUN-1998; 98US-0088811P.
 PR 10-JUN-1998; 98US-0088824P.
 PR 10-JUN-1998; 98US-0088825P.
 PR 10-JUN-1998; 98US-0088826P.
 PR 11-JUN-1998; 98US-0088861P.
 PR 11-JUN-1998; 98US-0088863P.
 PR 11-JUN-1998; 98US-0088876P.
 PR 12-JUN-1998; 98US-0089090P.
 PR 12-JUN-1998; 98US-0089105P.
 PR 16-JUN-1998; 98US-0089512P.
 PR 16-JUN-1998; 98US-0089514P.
 PR 17-JUN-1998; 98US-0089538P.
 PR 17-JUN-1998; 98US-0089598P.
 PR 17-JUN-1998; 98US-0089653P.
 PR 18-JUN-1998; 98US-0089908P.
 PR 19-JUN-1998; 98US-0089952P.
 PR 22-JUN-1998; 98US-0090246P.
 PR 22-JUN-1998; 98US-0090252P.
 PR 22-JUN-1998; 98US-0090254P.
 PR 24-JUN-1998; 98US-0090429P.
 PR 24-JUN-1998; 98US-0090435P.
 PR 24-JUN-1998; 98US-0090444P.
 PR 24-JUN-1998; 98US-0090461P.
 PR 24-JUN-1998; 98US-0090535P.
 PR 24-JUN-1998; 98US-0090540P.
 PR 25-JUN-1998; 98US-0090676P.
 PR 25-JUN-1998; 98US-0090678P.
 PR 25-JUN-1998; 98US-0090688P.
 PR 25-JUN-1998; 98US-0090690P.
 PR 25-JUN-1998; 98US-0090694P.
 PR 25-JUN-1998; 98US-0090695P.
 PR 25-JUN-1998; 98US-0090696P.
 PR 26-JUN-1998; 98US-00105413.
 PR 26-JUN-1998; 98US-0090862P.
 PR 26-JUN-1998; 98US-0090863P.
 PR 26-JUN-1998; 98US-0091010P.
 PR 01-JUL-1998; 98US-0091359P.
 PR 01-JUL-1998; 98US-0091544P.
 PR 02-JUL-1998; 98US-0091478P.
 PR 02-JUL-1998; 98US-0091486P.
 PR 02-JUL-1998; 98US-0091626P.
 PR 02-JUL-1998; 98US-0091628P.
 PR 02-JUL-1998; 98US-0091632P.
 PR 04-AUG-1998; 98US-0094006P.
 PR 04-AUG-1998; 98US-0095283P.
 PR 10-AUG-1998; 98US-0095998P.
 PR 10-AUG-1998; 98US-0096012P.
 PR 10-AUG-1998; 98US-009612P.
 PR 17-AUG-1998; 98US-0096757P.
 PR 17-AUG-1998; 98US-0096766P.
 PR 17-AUG-1998; 98US-0096867P.
 PR 17-AUG-1998; 98US-0096891P.
 PR 17-AUG-1998; 98US-0096897P.
 PR 18-AUG-1998; 98US-0096949P.
 PR 18-AUG-1998; 98US-0096959P.
 PR 18-AUG-1998; 98US-0097022P.
 PR 26-AUG-1998; 98US-0097952P.
 PR 26-AUG-1998; 98US-0097954P.
 PR 26-AUG-1998; 98US-0097955P.

CC The invention relates to an isolated cytoplasmic, nuclear, membrane bound
 CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a
 CC mature form, or a variant of the mature form of NOVX. Also included are a
 CC polynucleotide encoding NOVX (or its complement), a vector comprising the
 CC polynucleotide, a cell comprising the vector, an anti-NOVX antibody,
 CC determining the presence of NOVX in a sample using the antibody,
 CC which binds to NOVX polynucleotide, identifying an agent which binds to
 CC NOVX (including modulators of NOVX), NOVX, the polynucleotide and the
 CC antibody are useful for diagnosing, treating or preventing a NOVX-
 CC associated disorder selected from cardiomyopathy, atherosclerosis,
 CC diabetes, a disorder related to cell signal processing and metabolic
 CC pathway modulation, inflammation, autoimmune disorders, scleroderma,
 CC transplantation, allergies, systemic lupus erythematosus, haemophilia,
 CC graft versus host disease, Alzheimer's disease, stroke, Leech-Nyhan
 CC syndrome, periodontitis, pancreatitis, musculoskeletal disorders,
 CC Parkinson's disease, Huntington's disease, behavioural disorders,
 CC neurodegenerative and neuropsychiatric disorders, hypertension, wound
 CC healing, obesity, growth and reproductive disorders, lung diseases and
 CC many other diseases and disorders listed in the specification. NOVX, the
 CC polynucleotide and the antibody are useful in screening assays, detection
 CC assays (e.g., chromosomal mapping, tissue typing, forensic biology),
 CC predictive medicine (e.g., diagnostic assays, prognostic assays,
 CC monitoring clinical trials and pharmacogenomic), and in methods of
 CC treatment (e.g., therapeutic and prophylactic). NOVX is useful as
 CC immunogen to produce antibodies immunospecific for NOVX, as vaccines to
 CC screen for potential agonist and antagonist compounds, and as bait
 CC protein in a two-hybrid or three-hybrid assay. The polynucleotide is
 CC useful in gene therapy, to express NOVX, to detect NOVX mRNA or a genetic
 CC lesion in a NOVX gene, and to modulate NOVX activity. The vector is
 CC useful for producing non-human transgenic animals. The antibody is useful
 CC for isolating, and purifying NOVX and to monitor protein levels in tissue
 CC as part of a clinical testing procedure. The present sequence represents
 CC a NOVX protein
 XX
 SQ Sequence 1329 AA;

Query Match 43.7%; Score 2517; DB 5; Length 1329;
 Best Local Similarity 42.1%; Pred. No. 2.4e-105;
 Matches 571; Conservative 0; Mismatches 298; Indels 486; Gaps 86;

QY 1 GGGGCCCATTCGT-TTCAGCCATCGCC-----RAGATCATGAAGTCGCCATGGC 53
 Db 9 GCTGCCCTTCAGTCTCTGCACACTCTGGGCAACGACACCCGAT-CCGCTGTATC 67
 QY 54 -----AGCACC-----GCCACCGCC--GCCGCGGGCC--- 78
 Db 68 ACAACCGAGCCCTGTGGAGGGTGATGAGCAGCGGGGCATCTCTGGCCGAGAGCCTCA 127
 QY 79 -CCA--GCTGC-----GGCTGAA-----GGCCGCAACACAG-CGAGC 113
 Db 128 TCACGACTGACCTTATCACCAGGTGGCCCGCACTCTGGCTGGATCAGAGCTACC 187
 QY 114 GTTGGGGC-GAGGTGG---TCCGCTGTCTGT-----CTGAGCAGAGCGTGGCC---ATC 161
 Db 188 AGTCCTGCTGCGATPACCTTACC-TCACTGCCCCCTGGGCGGGG-GTCCCGGGGCAC 245
 QY 162 TCGCGCTGCGGGCGCGGGCGCGCTGCTGCTGCTGCTGGACGAGCAGG---T 217
 Db 246 CCGAGCCTCC-----GCCCGGTGTACCGTGCCTGGCC--GTTGG--GAGC--GAGGGGACT 295
 QY 218 AAACGTGTGTG-CTCT-----ACGATCATGAACG-----GCTGTATC 251
 Db 296 ACTCCCACTGTCTCTACACCAACGACATCACCAGGGTGCTGTACACCTTGCTGTATGC 355
 QY 252 TC-----ACGCTT-CAAGAGTGGT---GCCACCC--CTGC-----CCACG---AA 289
 Db 356 CATCATAGCTTCCATGCGTGTACCTTACCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 415
 QY 290 CCGCAAG-----GTGAGCAAGGTGG---AGATCTCCAGAC----- 323
 Db 416 CCGGTAGCTTTTTCAGACATGATGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 475

RESULT 7
 ABU8254
 ID ABU8254 standard; protein; 1300 AA.
 XX
 AC ABU8254;
 XX
 DT 07-JUL-2003 (first entry)
 XX
 DE Novel human secreted and transmembrane PRO polypeptide #2.
 XX
 KW Human; secreted and transmembrane protein; PRO; gene therapy;

QY 324 ---GTCA--TCGACTACATCAGGACCTTTCAGTTGGAGC-----TCAACT-----CGGAT 369
 Db 476 TGGTATATGTGACACAGATCAAGAGCT--GGTAGGGTGATGGTGACATCGCCAGCAA 533
 QY 370 CC-GAAGTTG--GGAC-----CCCCG---GGGGCGGAGG---GCTG 401
 Db 534 CCTGATGCTGGTGGAGAGCAGCCTGCTGTGGTGGGCCAGCGGAGGACAGAGCCTGCAG 593
 QY 402 CCG-----GTCC-----GGGTCGCTCAGCACC----- 425
 Db 594 CCGCATGCTGGTGGCTCGCTGGAGCGCATTTGGGGGGCGGCTCAGCCCATCGCCAGCA 653
 QY 426 ---CTCAAG--GCGAG-----ATCAG--CGCCCTG--A 450
 Db 654 CATCTAGTGAATCCGAGGAACCTGGCATTTGGAGGCTTACCTCATCAAGCCGACAGCTA 713
 QY 451 C---GGCCGAGGGCGCA-TGCTGTCTCGCGGA-----CGATCGCA---TC 488
 Db 714 CTTGGGCTGACCTGACAGCTTCCAGAGAGGAGGAGGGGTGCGGGGACACAGGCC 773
 QY 489 TTGTGTGCTG--AAGGCTTCCC-CCAGGACCGGCGGACCCAGCCATCCAGGG--- 542
 Db 774 AGGAAGCCCTGGCCAGAACCCCCACCTGAGCCCG--AGCCCCCAGCTGACCGAGCTC 831
 QY 543 -----GCA-----AGAGGAATTACG-TGCTCTGTGGTCTTCCCCAAC----- 579
 Db 832 CGCTTCGCTGACACCGGAGGCCCATTTCTCTGTCTGCTTCCATCAGAAC 891
 QY 580 -----GC-----GCTTC--GCGGAT-----CTGAGGAGAAACAGACCGAT--CGCGGGCC 622
 Db 892 AGCGTGGCCCTGGCCCTCCATCCAGCTGCGCCCGAGCTTATCTCATCCCTTCCGCTGCC 951
 QY 623 -----ACTCGGCCCT-TAACT----- 637
 Db 952 CTGGCTCCCCGGTGGCCCCCAGACTGCACCTGCAACTCTGCTTCCGAAATGCCGC 1011
 QY 538 -----GCATCCAG--CTGGGGCTGAGGCTGAGGCACTGGCGAGG 675
 Db 1012 CTCTTCACAGCCACAGCAACGCTCCCGCCCTGGAGCT--GCTG-GGC-CTGCAAGA 1066
 QY 676 AGAG-GGCGC-TCTCTCTGCAACACTAC-TAGTCACACAGAGACTTTAGGGGGTGGGATT 732
 Db 1067 GGCGTGGCGTGGCCACCCCGTCTCTTCGAGGAACAGTGGCTCT--GGCGTGGGAA 1124
 QY 733 CC--AC-----TCG-TGTGTTCTATTTTTTGAAGAGAGACATTTAA 773
 Db 1125 CCGTACAGAGTCAGTGGCCGCTTTCGCTGCGGCACTGGGCTGAGGAGCGCAACCTGTGGC 1184
 QY 774 AAATGCT-----CAC-----GTTTGGTGTCT--TCTCAGATTTCTGAGG 810
 Db 1185 CGCTTGTGGAGCCAGGAGGGGCCCGGGAGGCTGGGGCTGGACCTCGAGGGCTG--- 1241
 QY 811 AAATGCTTTGTTTGTATTTAATGATGATCAGG-ACTGAGAATATTTTACAATAG 869
 Db 1242 --CCAGCTCGCTCCAGCCAGCCCAATG-TCAGCGCCCTGCATGCCAG-----CACTTG 1293
 QY 870 TTCTGTGGGCTGTTTTTTTGTATTATAACAATA 904
 Db 1294 GGCAATGCGCGCTGCTCATGGAGCTGAGCCCTGA 1328

Db	585	ANGTCTTGGGCTTCCCGCCAGAGAGAGAGACCGCGCAAGTGTCTGTCACACGAGGAG	644
Qy	129	GTGCGCTGTCTGTAGACAGAGC-----GTGGCCATC-TGCG-----CTGC---170	
Db	645	GTGC-CGG--TGCCCGAGTGGGCCCCCGGGAGGAGCTCTGTCGCGGTCTCTGCGCAGCTCG	701
Qy	171	-----CGGGCGCGC--GGCGCGC--CTGCTTGGC--CTGCT-----202	
Db	702	GTCAACTACAACACCGTCTGCTGCTGTTTTCGAGCCGCTGCCACCTTCGGTTCTCTG	761
Qy	203	GGACG-AGCAGCAGG-TAAACGTGT-GCTCTACGACATGAACCGCTTTACTCA-----254	
Db	762	GAGCGCTACGGCGGCTCTCCGAGCTGGCGCGCGCGACGACCTGCCGTACCATCTC	821
Qy	255	-----CG-CCT-----CAAGAGCTG-----GTGCCACG--CTGCCCGAG--AACCG	292
Db	822	GGCTCGGACCTGGCGCGGTGTCTGTAGAGTGGGCCCCGGGTCAACCGTGGGGCGC	881
Qy	293	-----CAAGGT-----GAGCA-----AGGTGGAGAT--TCTCCAGC-ACGTTCATCGAC	332
Db	882	GGTGACGAGTCTGGCGCACTGCCTCTCGGTGGAGTGGAGTCCGCGACGGCCACCGC	941
Qy	333	TACATCAGG--GACC--TTGAG-----TTGAGCTG-----AACT-----CGGAT--369	
Db	942	GACACCATCTCGACCCGAGACGCGGATCTGGGCTTCGAGACCAACTTCGGCGGCTC	1001
Qy	370	-----AGTTGG-----GAGCCCGCGG--GG391	
Db	1002	GCGGAGATCGGTTGTCAAGCGACACGCTGATGCCAAACCGACCACTTCGACCTGG	1061
Qy	392	CGGAGGCTCGCGTCCGG-----CTCGCTCAGCACC-----TCACCGC434	
Db	1062	GAGGAGCGCGCGCGCGGACTGTGTCAACTCCACCGCTACCGCCAGCTGTCTCCGGC	1121
Qy	435	GAGATCAGCGCCTGAGC--GCGAGCGCGCATGGTT-----CCTCGCG--477	
Db	1122	-----AACGGGCGCGGATGAGAG--GGCAGCA-ACGTCTCTGCTGGGGGCGCAGCGCG	1176
Qy	478	-----ACGATCG-----CATTTGTGTG--CTGAAGCGC--TCCC-----510	
Db	1177	GTCTCGCGCGTTCGCCACCCAGCTGTGTGCGCGCGCGGCGCAATCCCGTCTCGTGG	1236
Qy	511	-----CCAGGACCGCGGACCCCGACGCATC-----CAGGGGC-----544	
Db	1237	TCTCCAGCGCGCGCAAGCGCGCATCTGCGTCCGATGGCGCGCGCGCTCATCGACC	1296
Qy	545	-----AAGAGGAATTACG-----TGCTCTGTGGTCTCCCCA-----A578	
Db	1297	GGGTCCGCGAGGACTACCGCTTCTGTCCGACGAGCGCACCGAATCCCGGGAGTGA	1356
Qy	579	CGCGCCTCG-----CCGATCTGAGGAGAACAGACCGATCGCGGCCACT--625	
Db	1357	AGCGCTTCGCGCAGCAGCATTCGGGAGCTACCGGAGCGGAGAGC--TCGACATGCTCTTC	1415
Qy	626	--GCGCCCTTAACTG--CATCC--AGCCTGGGGCTGAG-----GCT--GAGGCAC-667	
Db	1416	GAGCACCCCGCGGAGAGCTTCGGCGCTCGTCTAGCTGACCCGCAAGAGGACAC	1475
Qy	668	TTG-----CGAGGAGG-----GCGCTCTCTCTGCACACC--TACTAG---705	
Db	1476	GTGCTACCTTCGCGCTCGACGCGGTTTCGAGCAGCTTACGACACCGCTTACCTGTGG	1535
Qy	706	-----TCACCAGAGACTTTAGGGGGTGGGATTC734	
Db	1536	ATGTCCTGAAGCGCATCTGCGCAGCAGCTTCGCCAATTCGCGGAGCGTGGGAAGC	1595
Qy	735	ACTCGTGTCTTATTTTGAAG-----CAGACATTTAAATAATGTCAGTT788	
Db	1596	AACCG---GTTGTGTGTCAAGGCGAGATCCACCCGACGCT-----GTCCGCG--1640	
Qy	789	GGTCTCTCAGATTCT-GAGGAATTG-----CTTTGTATTGTATTATCAAA-836	

Db	1641	--TGCTACCCG-----CTGAGGAGTCCGCCAGCGGTCTACGAC--GTCCATCAAC	1691
Qy	837	-----TGATCAC-----CGAC---TGAGAATATTGTTTACAA	866
Db	1692	CTCCACACGAGGAGTCCGCTGCTCGCGTCCGCCGCGGAGGGCTCGGGTCCGG	1751
Qy	867	TA-----GTTCTGTGGGCTGTTTTTTTATTATAACAATAATTAGATGTTGTA	922
Db	1752	AACCCGAGTCCGGGAATGCCATCTTGCCG-CGATCAACCGCTCCG--GTTGCCGCC	1808
Qy	923	AAA925	
Db	1809	TGA1811	
RESULT 6			
AAU91279	AAU91279 standard; protein; 1329 AA.		
XX	AAU91279;		
DT	18-JUN-2002 (first entry)		
XX	Human NOV3a protein.		
XX	Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis; diabetes;		
KW	cell signal processing; metabolic pathway modulation; inflammation;		
KW	autoimmune disorder; scleroderma; transplantation; allergy;		
KW	systemic lupus erythematosus; haemophilia; Alzheimer's disease;		
KW	graft versus host disease; Lesch-Nyhan syndrome; periodontitis;		
KW	pancreatitis; musculoskeletal disorder; Parkinson's disease;		
KW	Huntington's disease; behavioural disorder; pain; obesity; wound healing;		
KW	neurodegenerative disorder; neuropsychiatric disorder; hypertension;		
KW	growth disorder; reproductive disorder; lung disease.		
XX	Homo sapiens.		
OS	WO200216600-A2.		
PN	28-FEB-2002.		
XX	27-AUG-2001; 2001WO-US026518.		
PF	25-AUG-2000; 2000US-0227800P.		
XX	25-AUG-2000; 2000US-0228205P.		
PR	25-AUG-2000; 2000US-0228324P.		
PR	30-AUG-2000; 2000US-0228997P.		
PR	30-AUG-2000; 2000US-0229185P.		
PR	01-SEP-2000; 2000US-0229780P.		
PR	01-SEP-2000; 2000US-0229848P.		
PR	01-SEP-2000; 2000US-0229850P.		
PR	22-JAN-2001; 2001US-0263337P.		
PR	31-JAN-2001; 2001US-0265518P.		
PR	15-MAR-2001; 2001US-0276451P.		
PR	27-MAR-2001; 2001US-0279196P.		
PR	24-AUG-2001; 2001US-00939398.		
XX	(CURA-) CURAGEN CORP.		
XX	Gerlach V, Macdougall JR, Smithson G, Stone DJ, Ellerman K;		
PI	Spytek KA, Zerkowen BD, Rastelli L, Verney CAM, Patturajan M;		
PI	Tchernev VT, Padigaru M, Taupier RJ;		
XX	WPI; 2002-292064/33.		
DR	N-PSDB; ABK55563.		
XX	New isolated cytoplasmic, nuclear, membrane bound and secreted		
PT	polypeptides, termed NOVX, useful for treating inflammation, autoimmune		
PT	disorders, hemophilia, Lesch-Nyhan syndrome, pancreatitis,		
PT	musculoskeletal disorders.		
XX	Claim 1; Page 32; 245pp; English.		
PS			

QY 217 TAAACGTG---CTGC-----TCTAC---GACATGAACGGCTGTACTCAAGCC--- 258
Db 458 AGATCAAGAGCTCTGAGATGTTCAAGTGTGACATGACATGATGACCTACCTCC 517
QY 259 ---TCAAGAGCTGTGCTGAC-CTGCCCCAGAACCG-CAAG-GTGAGC-AAGGTGGAG 311
Db 518 TCTTCAAGGAGT--GTGACCACTCCAAAC--GACCGTCTAGAGGGGCTGAGATCAG 573
QY 312 ATTCTCCAGC-ACGTC-----ATCGACTATACATAGG---GACCTTCAGTTGGAGTGA 360
Db 574 GAGTTCTCGCGGGCTGTGAAGCGCGGAGCTGGAGAGATCTTCCAT-----CAGT 628
QY 361 ACTCGG-----AATCG-AGTTTGGACCCCGGGG--GCCAGGG--CTGCCGTCC--- 408
Db 629 ACTCGGGCAGAGACCGCTGCTGAGTGGCCCTGAGCTGCTGAGTTCCTGGAGACCAAG 688
QY 409 ---GGGCTCCGCTCAG---CACCTCAACGCGAGATCAG-----CGCCC 447
Db 689 GCGAGAGGGCGGCACACTGGCGCGGCGCCAGCAGCTCATTCAGACCTATGAGCTCAAG 748
QY 448 TGACGGCCGAGGGCGATCGTTCTCGGGAC---GATCGCATTTG-----TG 493
Db 749 AGACAGCC-AAGCAGCATAG---CTGATGACACTGGATGGCTTCATGATGTATCTGTG 804
QY 494 TCGCTGAAG-----CGCCTCCC-----CCAG-----GGACCGG 521
Db 805 TCGCCGAGGGGACTCGCTTGGACAAACCCACACAGTGTGTTCAGGACATGACACAG 864
QY 522 C-----GGACCCACG--CATC-----CGCCTCCC-----C 537
Db 865 CCCCTTGGCCACTTCTCTCTCCACACACACTATCTGACTGACTCCAGATC 924
QY 538 AGGGGCGAGAGAAATGATG--CTCTGT---GGTCT---CCC-----CCACGCG- 582
Db 925 GGGGGGGCC-AGCAGACCGAGCCCTATGTATAGGGCTTTGCCAGGATGCGCTGTGT 983
QY 583 ---GGATCT-----GAGGAGAACAGACCGATCGGGCCA----- 623
Db 1044 CACCTCAAGATTTCTTTCGGGACGTGGCCCAAG--CGTGC-GGCACCATCGCTTCAG 1101
QY 624 CTGGCCCTTA-ACTG-CATCCAG--CCTGGGG-----CTG-AGGTGAGGCA-CTGGCG 672
Db 1102 CTGTCCCTTACCTGTCTATCTTATCCCTGGAGAACCACTGCGGGCTGGAGCAGGCG- 1160
QY 673 AGGAGAGGGCGCTCCT-CTCTGCAC-----ACCTACTAG----- 705
Db 1161 -TGCCATGGCCCGCCACCTCTGCACCATCTGGGGGACATGCTGTGACACAGGCGCTGG 1219
QY 706 ---TCACGAG--ACTTTAGGG-----GGT--- 726
Db 1220 ACTCCCAAAATCCGAGGAGCTGCCATCCCAGAGCAGTGAAGGCCGGTCTCTGTGA 1279
QY 727 -GGGANT-----CCACT-----CGTGTGT-----TTCTAT----- 750
Db 1280 AGGGAAGAAGCTGCCGCTCTCGGAGCGAGATGGCGGGCTCTGTCCGATCGGGAG 1339
QY 751 -----TTTGTGAACACACATTTTAAATAATGTT-----CA-----CGTT 787
Db 1340 AGGGGAGGAGATGACGAGAGGAGAGAGAGAGGAGGCTGGAGGCTGAGCGCAGCGCGC 1399
QY 788 TGTGTCTTCTCAGATTTCTGAGGAAT---TGCTTTGTATGTATATTAATATGATCACC 844
Db 1400 TG--GCCAAGCAGATCTCCCGGAGTGTGCGGCCCTG-GCTGTGTACTGCCAGCCACCC 1456
QY 845 GACTGAGATATTTT-----TACAATAGTTCTG-TGGGGCTG-TTTTTT 889
Db 1457 GCCTGGGAGCCCTGACCCCTGCGCCCAACGCCCAACACCTGCGAGGTGAGTCTCCTCA 1516

QY 890 GTTA---TTAACAATAATTTAGATGTTGTAAGAAAAA 926
Db 1517 GCGAGCGCAAGCCAGAAACTCATTCGGGAGCGAGGAA 1556
RESULT 5
AAE37007
ID AAE37007 standard; protein; 1811 AA.
XX AAE37007;
AC AC
XX XX
DT 23-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX XX
DE Micromonospora carbonacea polyketide synthase (PKS) type I #13.
XX XX
KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme.
XX OS Micromonospora carbonacea.
XX XX
FH Key Location/Qualifiers
FT Misc-difference 452..1811
FT /note= "Encoded by GCC"
XX CA2391131-A1.
XX 19-NOV-2002.
XX PD 26-JUL-2002; 2002CA-02391131.
XX PF 26-JUL-2001; 2001US-0307629P.
XX PR (ECOP-) ECOPIA BIOSCIENCES INC.
XX PA Yang X, Staffa A, Farnet CM;
XX WPI; 2003-343556/33.
XX DR N-PSDE; AAD55823.
XX Novel isolated polypeptide involved in biosynthesis of macrolides by
XX microorganisms, useful for biosynthesis of macrolides by microorganisms,
XX preferably for biosynthesis of rosaramicin.
XX Claim 13; Page 183-189; 206pp; English.
CC The invention relates to genes and proteins involved in the biosynthesis
CC of macrolides by microorganisms. In particular it relates to the nucleic
CC acids forming the biosynthetic locus for rosaramicin (a 16-member
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC useful for the biosynthesis of macrolides by microorganisms. It allows
CC direct manipulation of macrolides and related chemical structures by
CC chemical engineering of the proteins involved in the biosynthesis of
CC rosaramicin. It is useful to catalyse certain biochemical reactions, in
CC vitro or in vivo, to direct or enhance the synthesis or modification of a
CC polyketide, polyketide substrate or its precursor. The present sequence
CC is M. carbonacea polyketide synthase (PKS) type I protein. (Updated on 23
CC -OCT-2003 to standardise OS field)
XX Sequence 1811 AA;
Query Match 44.4%; Score 2562; DB 6; Length 1811;
Best Local Similarity 41.4%; Pred. No. 3e-107;
Matches 572; Conservative 0; Mismatches 313; Indels 498; Gaps 86;
QY 5 CCCATTCTGTTTCAGCCAGTCGCCAAGAATC-----ATGAAGTCGC-CAG-----TG 51
Db 465 CCCAGCTGAACCGATCTGACCGATCTCCGCCAGGAGACCGACGCGCGGAGCTG 524
QY 52 CGACACCGCC-----ACCGCGCGCG--CGGGCCCCAGCTCGCGC-----TGAAG-- 96
Db 525 GCGGCGCTGCGCTGCGCTCTCTCTACCGGCGCGTGCACGAGGACGAGACGCGG 584
QY 97 -----CCGSCAAGA-----CAGCG--AG-CGCTGCGGCG-CAGGTG 128

Db 2068 CCAGCAGCCCTCGTCCAGCCTGGCTCTCCAGCGCTGCCCTCTGCCAGATGGACCT 2127
 Qy 785 G-----TTTGG-----TGCTTCAGATTTCTCAGAA-----A 813
 Db 2128 GAGCACCACCTGACGCGCTTTTGGACCGCCACCTCTCTGCTGAAGAGGACAC 2187
 Qy 814 TTGCTTTGATTTATATTA-CAATGATCAGCTG--AGAATATTGTTTACAATAGT 870
 Db 2188 TGGCTATGAGGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2247
 Qy 871 TCTGTTGGGCTCTTTTGTGTTTATTAACAATATTTAGA-TGGTGAATAAAA 924
 Db 2248 GCAGATGAGGACCTGTTGACATT-----CTATTGAGCGGAGAAATTTCA 2296

RESULT 3

ABG74190
 ID ABG74190 standard; protein; 3907 AA.
 AC ABG74190;
 XX
 DT 29-APR-2003 (first entry)
 DE Mouse myocardin associated protein.
 DE
 XX Mouse; myocardin; cardiomyocyte growth; ANF induction;
 KW cardiomyocyte differentiation; sarcomere assembly induction; ANF;
 KW cardiac hypertrophy; atrial natriuretic factor; fibroblast modulation;
 KW non-cardiomyocyte cell modulation; cardiomyocyte; heart disease;
 KW cardiomyopathy; myocardial infarction; hypertension; gene therapy;
 KW cardiac-specific transcription factor.
 XX
 OS Mus musculus.

XX US2002164735-A1.
 XX
 XX 07-NOV-2002.
 XX
 XX 21-DEC-2001; 2001US-00029217.
 XX
 XX 21-DEC-2000; 2000US-0257761P.
 XX
 XX (OLSO/) OLSON E N.
 XX (WANG/) WANG D.
 XX
 XX Olson EN, Wang D;
 XX WPI; 2003-247258/24.
 XX
 XX Novel isolated polynucleotide encoding human or murine myocardin 1
 XX polypeptide, useful for modulating phenotype of non-cardiomyocyte cell
 XX e.g., fibroblast, to include phenotypic functions of cardiomyocyte cell.
 XX
 XX Disclosure; Page 51-61; 104pp; English.

XX The invention relates to an isolated polynucleotide encoding myocardin
 XX polypeptide. The effects of myocardin in growth and/or all
 XX differentiation of cardiomyocytes was assessed by overexpressing
 XX myocardin in cardiomyocytes using adenoviral delivering system. The
 XX results showed that overexpression of myocardin in neonatal
 XX cardiomyocytes induced assembly of sarcomeres and expression of atrial
 XX natriuretic factor, ANF, markers of cardiac hypertrophy. An expression
 XX cassette containing the polynucleotide operably linked to a regulatory
 XX sequence is useful for modulating the phenotype of a non-cardiomyocyte
 XX cell e.g. fibroblast to include one or more phenotypic functions of a
 XX cardiomyocyte cell. The expression cassette is useful for generating a
 XX cardiomyocyte which involves introducing into a cardiac fibroblast the
 XX expression cassette comprising the polynucleotide and a promoter active
 XX in the fibroblast, where the promoter directs the expression of the
 XX polypeptide. The expression cassette further comprises a second
 XX polynucleotide encoding GATA4 under the control of a second promoter
 XX active in a cardiac fibroblast. The expression cassette further comprises
 XX a polyadenylation site and an immunological marker. An expression

CC cassette comprising the polynucleotide encoding myocardin protein or
 CC peptide and a promoter operable in eukaryotic cells is useful for
 CC treating a heart disease, including cardiomyopathy, such as myocardial
 CC infarction or hypertension. The present sequence represents the amino
 CC acid sequence of the mouse myocardin associated protein. Note: the
 CC protein sequence presented is not disclosed in the specification but is
 CC shown in the sequence listing
 XX
 SQ Sequence 3907 AA;

Query Match 45.9%; Score 2648; DB 6; Length 3907;
 Best Local Similarity 38.6%; Pred. No. 7.7e-111; Indels 650; Gaps 85;
 Matches 600; Conservative 0; Mismatches 305;

Qy 1 GGGGC-COATCT- GTTTCAGCCAGTGGCCAGAAATCATGAA-----AG-----TC 44
 Db 761 GGGGCACCCCATGACTCATCT-ACGCCAAG-ATCCTGAGCAGCAGCAGCTCTTCC 818
 Qy 45 GCCAGTGGCAGCAGCCGCCAC-----GCCGCCGGG-GCCCCAGCTGGCG-----CTGA 93
 Db 819 TCCAGCTGCAGATCTCTCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 878
 Qy 94 AGG--CCGGC-AAAGACAGCAGCGGTGC-----GGCGAGGTG-----GTGCGCT 135
 Db 879 CTGCCCGCCAAAGTCAGCAGCGGAGGCGCTTGGGAAGCAGCGGAGCCCCCAGTACGCA 938
 Qy 136 GTCTGTCTG-----AGCA-----GAGCCT-----GGCCAT-----CTC 163
 Db 939 GCCTCTCACTACCAATAGCAGCTCCAGCTCGGGGCCCTGGGCTGTGGGTGGAC 998
 Qy 164 G-----CGCTCCGGG-GCGCGGGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 208
 Db 999 GTCAAGACAGCAGCTCACTGACTGGCAAGCC-GGGAGC-CCTGCCGGCCAACTGGACGA 1056
 Qy 209 GCAGCAGGTAAAGTGCTG-----ACTCAGCC-----TCTAC--GACATGAAC-- 242
 Db 1057 CATGAAGGTGGCAGAGCTGAAGCAGGAGCTGAAGTTGCGATCACTGCTGTCTCGGGAC 1116
 Qy 243 -----GGCTGTT-----ACTCAGCC-----TCAAG-----GAG-CTGG 270
 Db 1117 CAAAACTGAGCTGATTGAGCGCCTTCGAGCCTATCAAGCAAAATCAGCCCTGTGCCAGG 1176
 Qy 271 TGCC-----ACCTGCC--CCAGAAC-----CCCAAGGTGAGCAAGGTGGAG----- 311
 Db 1177 AGCCCCCAAGGCCCTTGGCGCCACCTCTATCTGTGACAAAGGTGGCGAGGTGTGGTAGC 1236
 Qy 312 ATTCTC--CAGCAC-----GTGATC-----GACTACA--TCAGGAGCTTC 348
 Db 1237 CTCCCGCAGCGCCCGGCTGAGCAGCGGCGCAGCCCTGTGGCAGCAGGCTGTGCTCCAGC 1296
 Qy 349 --AGTTGGAGCTGAATCTCGGAATCC-----GAAAGTTGGGACCCCGGGGGGC 392
 Db 1297 TGAGGTGTTGGTGGCCACGCTGGCCAGCAGTGGGGTGGTGAAGTTGGCAGCAGCGGCTC 1356
 Qy 393 CGAG-----GGCTGC-----CGGTCCGGGTCCGCTCAGCAC----- 424
 Db 1357 CACGCCCGCGTGTCTCCACCCCTCGGAGCGCTCACTGCTCAGCAGCGGCGATGAAAA 1416
 Qy 425 -----CCTCAAC-----GGCGAGAT-----CAGC--GCCCT 448
 Db 1417 CTCACCCCGGGGACACCTTTGTGAGATGTTGATCATCACCTCTGACGACGCTGACCT 1476
 Qy 449 GACGGCC-----GAGGGG-----CATGC--GTTCCT----- 473
 Db 1477 GAGGCTCCGCACTGAGATCTCTGTAAGGAGGAGGGGCCCCCGGGCGGGTCTCTGTG 1536
 Qy 474 -----GCGGACGAT-----CGCATC----- 488
 Db 1537 CCTGAGCCTCGGGGGCGGGCGGAGCTAGAGGGGCGCGCAAGGACAGATGCTGACGA 1596
 Qy 489 -----TTGTGTCCTGAAGCGC--CTCCCCCAGGGACCG----- 520
 Db 1597 GAAGACAGCAGATCAGGCGCTGAGCGCATGTCTCCGCGAGACGACGACGCTGCTGA 1656

XX	21-DEC-2000; 2000US-0257716P.	999	GTCAAGACAGCAGCCTCACTGCTGCAAGCC-GGGAGC-CCTGCCGCGCAACCTGGACGA	1056
PR	(TEXA) UNIV TEXAS SYSTEM.	QY	GCAGCAGTAAAGCTGCTG	242
XX	Olson EN, Wang D;	DB	1057 CATGAAGTGGCAGAGCTGAAGTGGAGCTGATCACTGCTGCTCTGGGAC	1116
PI	WPI; 2002-732693/79.	QY	243 -----GGCTGTT-----ACTCAGCC--TCAAG-----GAG-CTGG	270
XX	New myocardin polypeptides and polynucleotides, useful for respecifying	DB	1117 CAAGAACTGAGCTGATGAGCGCTTCAAGCCTTCAAGACCAATCAGCCCTGTGCAGG	1176
PT	non-cardiac cells, stimulating cardiac tissue regeneration, and for	QY	271 TGCC-----ACCTGCC--CCAGAAC-----CCCAAGGTGAGCAAGGTGAG--	311
PT	treating cardiovascular disorders, such as myocardial infarction and	DB	1177 AGCCCCCAAGGCGCTGCGCCACCTCTATCTCTGCACAGGCTGGCGAGGTGGTGTAGC	1236
XX	hypertension.	QY	312 ATTCTC--CAGCAC-----GTGATC-----GATACA--TCAGGACCTTC	348
PS	Disclosure; Page 146-159; 175pp; English.	DB	1237 CTTCCAGCGCGCGGCTGAGCAGCGGCGCAGCCCTGTGGCAGCAGGCTGTGCTCAGC	1296
XX	The invention discloses an isolated polynucleotide encoding a myocardin	QY	349 --AGTTGAGCTGAATTCGGAATCC-----GAAGTTGGGACCCCGCGGGCG	392
CC	polypeptide. Myocardial infarction results in the loss of cardiomyocytes,	DB	1297 TGAGTGTGTGGCCACGCTGGCCAGCAGTGGGTGGTGAAGTTTGGCAGCAGCGCTC	1356
CC	which are post-mitotic cells and generally do not regenerate after birth.	QY	393 CGAG-----GGCTGC-----CGTCCGGGCTCGCTCAGCAC-----	424
CC	Transplanting fetal cardiomyocytes has limitations so identifying new	DB	1357 CACGCCCGCGCTGTCTCCACCCCTCGGAGCGCTCACTGCTCAGCACCGGCGATGAAA	1416
CC	regulators of cardiomyocyte growth and differentiation is an important	QY	425 --CCTCAAC-----GGCGAGAT-----CAGC-----CAGC--GCTCCT	448
CC	goal in the search for therapeutics to treat myocardial tissue damage.	DB	1417 CTCCACCCCGGGGACACCTTTGTGATGTGTGATCACCCTCTCAGCAGCTGACCT	1476
CC	The polynucleotides, polypeptides and methods of the invention can be	QY	449 GACGGC-----GAGGCG-----CATGC--GTTCTC-----	473
CC	used to modulate the phenotype of a non-cardiomyocyte cell to include one	DB	1477 GCAGGCTCGCCACTGTCAGATCTCTGTGAAGGAGGAGGCGCCCGCGCGGCTGTG	1536
CC	or more phenotypic functions of a cardiomyocyte cell, to generate a	QY	474 -----GGGACGAT-----CGCATC-----	488
CC	cardiomyocyte, which comprises introducing into a cardiac fibroblast the	DB	1537 CCTGAGCCTTGGGGGCGCGGAGCTAGAGGGGCGGCAAGGACAGATGTGCAGGA	1596
CC	myocardin polynucleotide and a promoter which is active in the	QY	489 -----TTGTGCTGTGAAGCC-----CTCCCGCAGGACG-----	520
CC	fibroblast, stimulate cardiac tissue regeneration which comprises	DB	1597 GAAAGACAAGCAGATCGAGCGCTGACGGCATGCTCCGCGCAGAACAGCAGCTGTGA	1656
CC	inhibiting the function of myocardin in a post-mitotic cardiomyocyte and	QY	521 GCGGACCCAGCCATCCAG--GGGCAAGAGAGAAATACGTCTGTGTGCTCTCCCC--A	577
CC	for screening for a modulator of myocardin expression. The nucleic acid	DB	1657 GCGG-CTCAAG--TGAGCTGAGCAGGAGAA--GCGAGCCAGCAGCCCGCCCGC	1710
CC	can also be used in gene therapy to treat a heart disease, including a	QY	578 ACGGCGCTCGC-CGGATC-----TGAGGAGAGACA-----AGACCGAT--	614
CC	cardiomyopathy, comprising administering a polynucleotide encoding a	DB	1711 CCGCGCCCGCTCGCACCCCGCTGAGCAGGAGAGAACAGCTTCTCAGCTGCCAGCTGAG	1770
CC	myocardin peptide or protein or an antisense nucleic acid. The	QY	615 -CGCGCGCCACTGCGCC--CTTAAC--TGCAAT--CCAGCCTGGG--	652
CC	polynucleotide and polypeptide can also be used for decreasing mortality	DB	1771 CCAGCAGCCCTGGGCGCGCTCACCATTTCAACCCAGCTGGCGCGCCAGCCACCAA	1830
CC	in a subject with heart failure, comprising inhibiting the function of	QY	653 -----CTGAGCTGAGGCACTGG-----CGAGGAG-----AGGCGC	684
CC	myocardin in post-mitotic cardiomyocytes, increasing the level of	DB	1831 CCACATAGACCTTGTCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1890
CC	myocardin in fibroblasts to generate cardiomyocytes, inhibiting the	QY	685 T-----CC--TCTCTGCACAC-----CTACTAG-----TCA-----C	709
CC	function of myocardin in post-mitotic cardiomyocytes or increasing the	DB	1891 CTTGACGCTGAGCCCGGAGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1950
CC	level of myocardin in fibroblasts to generate cardiomyocytes in the	QY	710 CAGAGACTT--TAGGGGTGG-----GATTCACCT-----CG	739
CC	subject. The compositions and methods of the present invention are useful	DB	1951 CAGCCTCATCAGGGGTGTGACCTCCACCCTCATCAGGACTCCACAGGACCCACCT	2010
CC	for respecifying non-cardiac cells, stimulating cardiac tissue	QY	740 TGT-----GTTTCTATTTTTTGAAGACAGACA-----	767
CC	regeneration, for treating cardiovascular disorders, such as myocardial	DB	2011 TGTCTCAGCGTGACCA--ATAAGAAATGCAGACAGCCCTGGCGCTGTCCAGTGGGAGCC	2067
CC	infarction and hypertension, and for screening compounds for various	QY	768 -----TTTTAAATATGTCAC-----	784
CC	abilities to interact and/or affect myocardin expression or function. The			
CC	sequence presented is the mouse myocardin associated protein, #1. Note:			
CC	This sequence is presented as a 3-letter coded protein sequence in the			
CC	specification, but when changed into the single letter code, forms a DNA			
CC	sequence identical to that given in AB55224. It, therefore, appears to			
CC	be the DNA sequence encoding human myocardin 2			
XX	Sequence 3907 AA;			
SQ	Query Match			
	Best Local Similarity 45.9%; Score 2648; DB 5; Length 3907;			
	Matches 600; Conservative 0; Mismatches 305; Indels 550; Gaps 85;			
QY	1 GGGGC-CAATCT--GTTTCAGCAGTGTGCCAAGATCATGAA-----AG-----TC	44		
DB	761 GGGGACCCCGCATGACTATCT-ACGCCAAG-ATCCTGCAGCAGCAGCTCTTCC	818		
QY	45 GCGAGTGCAGCAGCGCCACC-----GCGCGCGCGG-GCCCGCAGCTGCGG-----CTGA	93		
DB	819 TCCAGCTGCAGATCTTCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	878		
QY	94 AGG--CCGGC-AAGACAGCAGCGGTGC-----GGCGAGGTG-----GTGCGCT	135		
DB	879 CTGCCCCCGCAAGATCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	938		
QY	136 GTCTGTCTG-----AGCA-----GAGCGT-----GCCCAT-----CTC	163		
DB	939 GCCTCTCCACTACCAATAGCAGCTCCAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCG	998		
QY	164 G-----CGCTGCCGGG-GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	208		

CC glomerulonephritis, Hashimoto's thyroiditis, ischaemic injury, multiple
 CC myeloma, multiple sclerosis, osteoporosis, Parkinson's disease, psoriasis
 CC and vasculitis. Disorders may be treated with the novel composition using
 CC gene therapy. This polynucleotide sequence represents a TALL-1 related
 CC DNA sequence of the invention
 XX
 SQ Sequence 7285 AA;

Query Match 46.5%; Score 2681.5; DB 6; Length 7285;
 Best Local Similarity 37.1%; Pred. No. 3.9e-112;
 Matches 620; Conservative 0; Mismatches 294; Indels 757; Gaps 83;

QY 1 GGGGCCCATCTTCT--GTTTCAGCCAGTCG---CCAGAA-----33
 DB 2671 GATCCCGCGTGGAGGATCATCAGCGCGCGTCCGGAACGATCCGAGCCCAACC 2730
 QY 34 --TCATGAAGTCCGAGTGCAGCAGCCGAC-----CGCGC-----70
 DB 2731 TTTTCATAGAGCGCGGTGGAATCTCGTATGTCAGGTTGGCGCTCGTTGCT 2790
 QY 71 CG-----CGGGCCCCAGCTGCGCT--GAAGGC--CGGCAAGACAGCA-----GC 113
 DB 2791 CGGTCAATTGAAACCCAGAGTCCGCTCAGAGAACTGTCAGAGAGCGATAGAGC 2850
 QY 114 GGTG-----CGGCGAGGTGTGTG-----GCTGTCTG-----TC 142
 DB 2851 GATCGCTCGAATCGGAGCGCGATACCGTAAAGCAGAGGAGCGTCAAGCCATTC 2910
 QY 143 TGAGCAGAGC-----GTGGCCATCTC-----GCG-----166
 DB 2911 GCCGCGAGCTTTCAGCAATATCACGGGTAGCCAGCTATGTCTGTATGCGGTCCG 2970
 QY 167 -----CTGCCG-----173
 DB 2971 CACACCCAGCGGCCACAGTCGATGAATCCAGAAAGCGGCCATTTTCCACCATGATTT 3030
 QY 174 -----GGCGCCGG-----GGCGCCCT-190
 DB 3031 CGGCAAGCAGCATCGCCATGAGTCACGACGAGATCCTCGCTCGGCGATCGGCGCTT 3090
 QY 191 --GCTGCG-----CTGCTGG--ACGAGCAGAG-----TAAAGTGTGTCT-----230
 DB 3091 GAGCTGCGCAACAGTTTGGCTGGCGGAGCCCTGATGTCTTCTGTCAGATCATCTG 3150
 QY 231 TACGACATGAACGGCT-----GT--TACTCAGC-----CT-----259
 DB 3151 ATGACACAGCGGCTTCCATCCGAGTACGTGTCTCGTCTGATCGATGATGTTTGGTGG 3210
 QY 260 -----CAAGAGGTGG-----TGC-----CCACCTGCGCCAGAAC-----291
 DB 3211 GTCGAATGGCAGGTAGCCGATCAAGCGTATGACGCCGCCGATTCGATCAG--CCATG 3268
 QY 292 -----GCAAGGTGAGCAAGTGGAGATTTCT---CCAGCAGCTCA 327
 DB 3269 ATGGATATTTCTCGGAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGCACTTCG 3328
 QY 328 TGGACTA--CATCAGGACCTTC-----AGTTGG-----AGTGAATCGCA 367
 DB 3329 CCCAATAGACCCAGTCCCTTCCCGTTTCAGTGAACACGTGACAGCAGCTGCGCAAGGA 3388
 QY 368 A--TCCGAAGTTGGACCCCGGGCGCGCGAGGGTGC-----402
 DB 3389 ACGCCCGTCTGGCGCAGCAGCAGATAGCCGC--GCTGCTCTGCTGCAATTCATTTCAGGA 3446
 QY 403 -----CGGT-----CCGGGCTCGCTCAGC-----ACC-----CT 427
 DB 3447 CACCGGACAGTCTGCTTTCGAAAAAGAACCCGGCGCCCTCGCTGACAGCCGGAACA 3506
 QY 428 CAACGGCG--AGATCAGC-----GCCCTGAC--GGCCG-----456
 DB 3507 CGGGCGCATCAGAGCAGCGGATTTGCTGTTGTCGCCAGTCAATAGCCGAATAGCTCTCCA 3566
 QY 457 ---AGCGGCGATGCTTCTCTCGGACGATCGCATCTGTGCTGCTGAAGCG-----CC 506

DB 3567 CCCAAGCGCGCGAGAACCTGGTGCAATC-CATCTTGT-TCAATCATCGCAAAAGATCC 3624
 QY 507 TCCCCCAG-----GGACCGCGC---GGACCCC--AGCCATCC-----AGGGGCAAG-A 548
 DB 3625 TCATCTGTCTTGTATCTGATCTGATCCCTCGCCATCAGATCCTTGGCGGCAAGAA 3684
 QY 549 GGAATFAG--TGCTCTGTGGTCTCCCCCAAC-----GGCGCTCG-----587
 DB 3685 AGCCATCCAGTTTACTTTTCAGGGCTTCCCAACCTTACCAGAGGGCGCCAGCTGGCAA 3744
 QY 588 --CCGATCTGAGGGAGAAC--AAGACCG-----ATCGGC-----GGCCACT 625
 DB 3745 TTCGGTTGCTTGTGTCATTAACCCGCCAGCTAGCTATCGCATGTAAGCCACT 3804
 QY 626 GC-----GCCCT--TAAT 637
 DB 3805 GCAAGCTAOCCTTCTCTTTGCGCTTTCGCTTTTCCCTTGTCCAGATAGCCAGTAGCT 3864
 QY 638 G-CATCCAGCTCGGGCTG-----AGGCT-----660
 DB 3865 GACATTCATCCGGGGTCAGCACCGTTTCTCGCGACTGGCTTTTCTACGTGTTCCGCTTCT 3924
 QY 661 ---GAGCACTGGCAGGAGAGGCG-----GCTCC--687
 DB 3925 TTAGCAGCCCTTCCGCCCTGAGTCTTGGCGCAGCGTGAAGCTACATATATGTATCGG 3984
 QY 688 -----TCTCTGCA-----CACCTAC--TAGTCAACAGAGAC-----TTT 719
 DB 3985 GCAATGCGTGAATATTCCTTTTGTCTCGACCATCAGGCACCTGAGTCTGCTCTTTT 4044
 QY 720 AGGGGTGGGATTC-----CACTC-----GTG-----TGTT 746
 DB 4045 C-----GTGACATTCAGTTCGCTCGCTCAGGCTTGGCAGTGAATGGGGTAAATGCA 4100
 QY 747 CTA-----TTTTTTGAA---AAGCAG-----ACATTTTAAAAAATGGTCACT 786
 DB 4101 CTACAGCGGCTTTTATGATTCATCAAGAGAACTACCCATAATACAGAAAGACCCGT 4160
 QY 787 TTGCTGCTTCTCA---GATTTCTGAGGAAATGCTTTGTATTG-TATATTAC-----AA 836
 DB 4161 CAGCGGCTTCTCAGGCGGTTTATGCGGGTCTGCTATGTGTGCTATCTGACTTTTTC 4220
 QY 837 TGATCA-----CCGACTCAGAAATATTTTCTTAC-----AATAGTTCTGTG---876
 DB 4221 TGTTCAGAGTTCCCTGCCCTCTGATTTTCCAGTCTGACCCTTCGGATTATCCCGTGACA 4280
 QY 877 GSGCTGTTTTTG--TTATTAAACAATAATTTAGATGGTGAATAAAAAA 925
 DB 4281 GGTCACTCAGCTGGCTAATGACCCAGTAAGGAGC--GGTATCATCAACA 4330

RESULT 2

ABG70822

ID ABG70822 standard; protein; 3907 AA.

XX AC ABG70822;

XX DT 16-DEC-2002 (first entry)

XX DD Mouse myocardin associated protein #1.

XX DE Mouse; myocardin; myocardial infarction; cardiomyocyte;
 XX KW post-mitotic cell; differentiation; therapeutic; gene therapy;
 XX KW heart disease; cardiomyopathy; mortality; heart failure; hypertension.
 XX OS Mus sp.

XX WO200260946-A2.

XX PD 08-AUG-2002.

XX PF 21-DEC-2001; 2001WO-US050606.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:10:55 ; Search time 66 Seconds
(without alignments)
3964.227 Million cell updates/sec

Title: X77956

Perfect score: 5766

Sequence: 1 GGGGCCCATTCGTTTCAGC.....TTAGATGCTGAAAAA 926

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681.5	46.5	7285	6	ABJ38280 PAMG21-RA
2	2648	45.9	3907	5	ABG70822 Mouse myo
3	2648	45.9	3907	6	ABG74190 Mouse myo
4	2596	45.0	2211	6	ABG72959 Novel hum
5	2562	44.4	1811	6	AAE37007 Micromono
6	2517	43.7	1329	5	AAU91279 Human NOV
7	2491	43.2	1300	6	ABU88254 Novel hum
8	2491	43.2	1300	6	ABU90133 Novel hum
9	2491	43.2	1300	6	ABU96435 Novel hum
10	2491	43.2	1300	6	ABU99044 Novel hum
11	2491	43.2	1300	6	ABU98259 Novel hum
12	2491	43.2	1300	6	ABU91965 Novel hum
13	2491	43.2	1300	6	ABU95269 Novel hum
14	2491	43.2	1300	6	ABO00408 Novel hum
15	2491	43.2	1300	6	ABU88959 Novel hum
16	2491	43.2	1300	6	ABO06455 Novel hum
17	2491	43.2	1300	6	ABU95515 Novel hum
18	2491	43.2	1300	6	ABU95205 Novel hum
19	2491	43.2	1300	6	ABU90753 Novel hum
20	2491	43.2	1300	6	ABU99315 Novel hum
21	2491	43.2	1300	6	ABU86189 Novel hum
22	2491	43.2	1300	6	ABU82044 Novel hum
23	2491	43.2	1300	6	ABU07905 Novel hum
24	2491	43.2	1300	6	ABU94225 Novel hum
25	2491	43.2	1300	6	ABO00098 Novel hum

26	2491	43.2	1300	6	ABU87109 Novel hum
27	2491	43.2	1300	6	ABU91350 Novel hum
28	2491	43.2	1300	6	ABU90443 Novel hum
29	2491	43.2	1300	6	ABU97034 Novel hum
30	2491	43.2	1300	6	ABO05230 Novel hum
31	2486.5	43.1	1679	4	AAU07343 1-aminocyc
32	2486	43.1	1089	6	ABM15834 Mycobacte
33	2445.5	42.4	1743	6	ABU88255 Novel hum
34	2445.5	42.4	1743	6	ABU90134 Novel hum
35	2445.5	42.4	1743	6	ABU96436 Novel hum
36	2445.5	42.4	1743	6	ABU99045 Novel hum
37	2445.5	42.4	1743	6	ABU98260 Novel hum
38	2445.5	42.4	1743	6	ABU91966 Novel hum
39	2445.5	42.4	1743	6	ABU85270 Novel hum
40	2445.5	42.4	1743	6	ABO00409 Novel hum
41	2445.5	42.4	1743	6	ABU88960 Novel hum
42	2445.5	42.4	1743	6	ABO06456 Novel hum
43	2445.5	42.4	1743	6	ABU95516 Novel hum
44	2445.5	42.4	1743	6	ABU95206 Novel hum
45	2445.5	42.4	1743	6	ABU90754 Novel hum

ALIGNMENTS

RESULT 1
ABJ38280
ID ABJ38280 standard; protein; 7285 AA.
XX AC ABJ38280;

DT 12-JUN-2003 (first entry)

DE PAMG21-RANK-Fc vector protein SEQ ID NO 28.

XX TALL-1-binding protein; TALL-1; B-cell-mediated autoimmune disease;
KW systemic lupus erythematosus; B-cell-mediated cancer; lymphoma;
KW inflammation; rheumatoid arthritis; acute pancreatitis; atherosclerosis;
KW Alzheimer's disease; asthma; cachexia; cirrhosis; diabetes; osteoporosis;
KW glomerulonephritis; Hashimoto's thyroiditis; ischaemic injury; psoriasis;
KW multiple myeloma; multiple sclerosis; Parkinson's disease; vasculitis;
KW gene therapy; ds.

XX Unidentified.

PN WO200292620-A2.

XX 21-NOV-2002.

XX 13-MAY-2002; 2002WO-US015273.

XX 11-MAY-2001; 2001US-0290196P.

XX (AMGE-) AMGEN INC.

XX Min H, Heu H;

XX WPI; 2003-156719/15.

XX New TALL-1-binding polypeptide, useful for modulating the activity of
TALL-1 and in treating, preventing or diagnosing a B-cell-mediated
autoimmune diseases, cancers or lymphomas.

XX Disclosure; Fig 4; 236pp; English.

XX The invention relates to a novel TALL-1-binding polypeptide comprising a
defined sequence in the specification. The composition is useful in
modulating the activity of TALL-1, and in treating, preventing,
ameliorating, diagnosing or prognosing a B-cell-mediated autoimmune
disease (e.g. systemic lupus erythematosus) or B-cell-mediated cancer or
lymphoma. The composition may also be used in treating inflammations
(e.g. rheumatoid arthritis), acute pancreatitis, Alzheimer's disease,
asthma, atherosclerosis, cachexia, cirrhosis, diabetes,

Db	2491	NNTAKRLLTCADIONGTSTVCVVALSTCVNSGTACIPKANGCTYTNKIACNSGGLDGI	2550
Qy	1191	---TTTATATAAAGT	1202
Db	2551	CVFTQSTATGAVAGT	2565

Search completed: May 7, 2004, 15:10:02
 Job time : 74 secs

[illegible]

RESULT 14

Q8VIZ1 PRELIMINARY; PRT; 1384 AA.
 AC Q8VIZ1; 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE PE_PGRS family protein.
 GN Mycobacterium tuberculosis.
 OS Bacteria; Actinobacteria; Actinobacteriales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN 111_1773;
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., PeBooy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB007163; AKA47970.1; -;
 DR TIGR; MT3612; -;
 DR InterPro; IPR002952; Eggshell.
 DR InterPro; IPR000084; PE region.
 DR InterPro; IPR002173; PFKB.
 DR Pfam; PF00934; PE; 1.
 DR PRINTS; PR01228; EGGSHLL.
 DR PROSITE; PS00583; PFKB KINASES 1; 2.
 SQ SEQUENCE 1384 AA; 110839 MW; 124AC8AAADD5EBD CRC64;

Query Match 11.5%; Score 869.5; DB 16; Length 1384;
 Best Local Similarity 30.4%; Pred. No. 2.2e-43;
 Matches 418; Conservative 23; Mismatches 612; Indels 321; Gaps 54;

QY 1 GATCTGGGTGCTCCAGCAAAAGCAAAATCTCGAAGTTAATGGTTTTCAGTGAATTTT 60
 DB 147 GAGGSGNGSAGLWNGNGAGGAG- - - - -GSGGAGGNGGNGWLFAGGTGGTGGT 200
 QY 61 AAATCTCTCTGGCGAGA- - - - -GGCCCGCTCTCCCGGTATCAGCCCTTCCTCATCTT 117
 DB 201 AFG- - - - -AMGGTGGNGGALLTG- - - - -GGLGAGGNGGTGGTGGTGGNG 244
 QY 118 TGAATCCGGCTCCGGCTCTTCGGCGTCA-GACCAGCCGAGGAGAGCCTGTTTGA- - 174
 DB 245 GNCALLIAGGVG--GAGG- - - - -IGGQGTGAGGAGAGGTGNGGAGGLFWNGDGGAGG 298
 QY 175 - - - - -ATTAAAGCG- - - - -GCTGTGAACGCCAGCGCCGCGCGCGCGCGGCGC 218
 DB 299 QGGDGAAGDAASAGCTGGKGGQGGDGTGGAGAGPVLFGHGGAGMGQGGTGGMG- - 356
 QY 219 GAGCGCGGCAATTT--TGAATAAAGAGCGGTGCGCTTCAGGACGCTCTATAAGTGACCG 276
 DB 357 GAGGDTTVAAGTGGEGTGAAGAGG-AAGARGALTSGLAGG- - - - -VGAGGTG--G 408
 QY 277 CCGCGCGAGCGTGGCGG- - - - -GTTGAGGTCACTAGCGGACTCTTTTGGTTTCTTTC 335
 DB 409 TGGTGGNGADAADVGVFGANGDPGAGGK- - - - -GNGGGGGA- - - - -AVTGG- - - 451
 QY 336 TCTTTGGGCACTCTGCACTCACTCCCGAGCATGAAGGGCGCTGAGCCCGGTGCGGGCT 395
 DB 452 - - - - -VAGDGTGGKGTGA-GGAGNDAGSTNGPGKGGDGGIGGAGGAGGAAGTNGGHA 507
 QY 396 GCTACGA-GCGGTGTGTCGCTG- - - - -TCGGA- - - - -ACCACTGCGGCATGCC 442
 DB 508 GNTGDDGDDGTGGNGNGTGGVNGADNTLNPTDGGAGGPGGAGGAGGAAGGCGGTGG- - 565
 QY 443 CCGGGCCGAGGAAGGG- - - - -CCCGCAGCTGAGGAGCGCGCTGAGCT- - - 485

DB 566 TGGNGNGNGNGNGNGNGNGNGNAGNNSNAPVGGEGGAGGDDGAGGAGGAANGGTAG 625
 QY 486 - - - - -TGCTGAGCAGCATGAACCACTGCT- - - - -ACTCCCGCTTGGCGGAACCTGGTAC 533
 DB 626 SQGTGGVGGDGGAGNGGGGKAGTNSGNFVGDGEAGFSGGAGNGGVGGAAGANGGTGG 685
 QY 534 CCGGAGTCCCG--AGAGGCACTCAGCTTAGCCA- - - - -GGTGAATCTCAGAGCG 583
 DB 586 SGGNGDGGAGGIGGAGGNGIPGTGTPAGGTGAKGGDGGDGGAGGAGNAGGAGNGA 745
 QY 584 - - - - -GTCACTGACTATCTT- - - - -CGACCTGCACTAGTCTTGGCCGAGCAGCCCTG 635
 DB 746 GGGGAGGAGGAGGAGGAGNAPVGDGVGKAPHGNDAGSGDGGKGGGGSGGSGGSGGAPTG 805
 QY 636 GACCCCTGATGGCCCCCACCCTTCCCATCCAGACACCGAGCTCGCTCCGGAACCTGCA 695
 DB 806 GG- - - - -AGTGG- - - - -SGHAGKGGAGG- - - - -GAGGTITIVP 836
 QY 696 TCTCAACGACAAAGAGG- - - - -TTTGGCACTGACTCGGCGGTGCTCTGACACCTCC 750
 DB 837 GNGNAGDGGNAGAGGNGSGDFGNTTSG- - - - -ASGGGNGGAGT- - - - -A 880
 QY 751 AGAACGCACTGCTGGCGCCGCTTCTGCTGGGACCCCG- - - - -GGAACCTC 797
 DB 881 GSGGAGGTGCTGLSGNGNGNGNGGNGGNGGNGAGTGAQFVATSLPTPNAGGNGG 940
 QY 798 TCTGCGCGAAGCCGAGCGGAG- - - - -GGATGGGCCCC- - - - -AAC 834
 DB 941 TGSNGGAPGAPGAPGPTTGGNAGSOGIGDGGNGDGGGADAVNVFMEFPQAT 1000
 QY 835 TTCGCCCTGCCACTTGACTTCAACAA- - - - -ATCCCTTCCTG 872
 DB 1001 GTAG--SAGDPTGGNGGPGTPGSPWVAPPPPIITQVQGGDGGAGGTGSTNANDGTATG 1058
 QY 873 GAGACTAACTGCTCTCAGGAGCAAG- - - - -ACTGTGAATGT- - - - -AGCC 919
 DB 1059 KGK- - - - -EGVGSILGGPGNGGTGGNAGSATGNGVANAGNGKGGDGGGQFGAGN 1111
 QY 920 TGAAGACCCAGAGTACTCTGGCC- - - - -ACCAGCTGGCGGACGCTCACCTTGCTCCACCCC 977
 DB 1112 GGAGSVTDSAGSTAGNGNGNATNGTIAGPAGNGSAGKG--GDGNG- - - - -1161
 QY 978 ACCCCCAAGTCTAA--GGTCTTTTACAGCGGTGAGGTGGAAGAGGTGGCTCTCTCC 1036
 DB 1162 - - - - -IAAGATGAGNGNGNGNDGAVNAGTGGGNG--GNAGGGANGGDDG- - - - -1209
 QY 1037 AAATATGCCAAGCGCGCGCAGAGCTGCTTCTGCTCTCTCTTGGAGAAAGTCTGTT 1096
 DB 1210 - - - - -AGAGGAGGKGGKIDGGFGDGGNGGNGGTGAGGNGGNGTGGVSV 1258
 QY 1097 GCCCTGATTTATGAATCTATATAGTATATAGTTTGTACCTTTTACAGAGG 1156
 DB 1259 GAAGDGGNGGTGGFAGGTAGNGSG--GTGAGDGGTGGDGGNGGTGTAGGGTGG 1317
 QY 1157 TGACTTT--CTGTAACA- - - - -ATCGCATGTAT--ATTAACTTTTATAAAAGT 1202
 DB 1318 NGGASGAGGAGGTGGFAGNGNAGNGGTGGASDGDNGNAGSGATGCTGGNGGT 1371

RESULT 15

Q94710 PRELIMINARY; PRT; 2717 AA.
 ID Q94710
 AC Q94710;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE 51A surface protein.
 OS Paramacium tetraurelia
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Penicillida;
 OC Paramacium.
 OX NCBI_TaxID=5888;
 RN 11

Db 431 TTCAATLLEAACVKGSTGCFWDTASSCKDKICVNAPATNTTHDLCOAFLNTCTVNST 490
Qy 162 AAGC-----CTGTTTGA-----ATTTAAGC---GGGCT- 187
Db 491 SAGCEKTCENSILVLAICDKDTSNRACIKWKYKQCVLASSATTHADQTVHSSCTL 550
Qy 188 ---GTG---AAGGCCAGGGCGGGCGGGCGGGCGGAG-----GC 223
Db 551 SNQGTGCVPLPLKCEAITLERACN-MKANGQPCG-WNGSQICDKACSTASKTATTTQC 607
Qy 224 GG-----GC----- 227
Db 608 QGHISTCVANNPATVNGSLITGQCQDLPTTCAARKSTENCEISRAQFTCLWVASTSCV 667
Qy 228 ---CATTTTGAATAAGAGCGCTGCC---TTCCAGGAGGCG---TCTAT 267
Db 668 EXSCATASAGTTGALSAGGFTFSACQAYLNTCISNNADGCMWKPSSCSLVSNCQLG 727
Qy 268 AAGTGAC-----C-----GC-----CGGGCGGAGC---GTGC--- 291
Db 728 SKASGDCVWNGSCVDRTCANITQTTHNSCOSTFNQCTVNNGSTACQPLATACTSYSTQE 787
Qy 292 GC-----CGGTG---CAGGTCA-----CT-----GTAGC- 313
Db 788 NCKFTSNKNCVWTGLACRNATCADAPDTASDTDECLAYTTPSETCTVYVYKGAQGC 847
Qy 314 --- 319
Db 848 PKSANCNVYMTQACHKTITNLTDNDCKWVDRVCVAISSFATGACTTFFKGTQTMCOGYR 907
Qy 320 ---CTTTTGGTTTT-CTTTCTCTT-TG-----GGGCAC---CTCTGGAC 355
Db 908 AGCNTGGATSSFACTLDCTLTGTGLTFADQALDSTCSVKDGTGCVIIOGACTGYGT 967
Qy 356 TCACTCCCGCAGCATGAAGCGCTGAGCC-----CG----- 385
Db 968 TAA---NCFRSSASGNAGYCALNGSSTCVTVSAAADCAFTGLTGLDHTKCOLYHTSCTA 1024
Qy 386 ---GTGC-----CGGGCTGTACAGCGGTGTGC-----TGCTGTGGA 423
Db 1025 LKDGTTGCQYKTTCSGYATNNCATSGQKCPDGTDLRFNTHNCASVTGTGLTTICATY 1084
Qy 424 ACCGAGTCTGCC---ATGCCCGGGCGGAGGAGGCGCGG-----C 465
Db 1085 DAGCIANVDGTAQCKLAUCDLVLTQKSCSTSAARATADKCAWSGTACLAIVTANIATHC 1144
Qy 466 AGCTGAG-GAGCCCGCTGAGTCTGAGC-----GACATGAAC---CACT 506
Db 1145 AVVTGTGLTDAICAAVNVDTANRAGTACQEQKATCDLYTTEATCSTSAARATADKCAWS 1204
Qy 507 GCTACTCCCGCTGC---GGG-----AACT---GGTAC-----C 534
Db 1205 G-TACLAVTVATHCAVKGTLTDTTCAAYNADCTANRAGTECQEQKATCDLYTTEATC 1263
Qy 535 ---CGGAGTCCCGAGAGGCACTCA-----GCTTAGC-----CAGGTGG--- 569
Db 1264 STSKAATADKCAWSGTACLAIBETANIATHCAVVGKGTGLTDTATCAAYNVDTANRAGTEC 1323
Qy 570 ---AAATCTACAGCGCTCA---TCGACTACATTC-----TCGA 603
Db 1324 QEQKATCDLYTTEATCSTSAARATADKCAWSGTACLAIVTVATHCAVVGKGTGLTDTCAA 1383
Qy 604 CCTCAGGTAGTTC-----TGGCCG---AGCCAGCCCGCTGAGC-----CC 640
Db 1384 YNADCTANRAGTECQEQKATCDLYTTEATCSTSKAARATADKCAWSGTACLAIVTVATHCA 1443
Qy 641 CCTGATGGCCCCCACTTCC---ATCCAGACAGC-----CGAGTCGCTC--- 683
Db 1444 YVTG-TGLVATCAAYNVGICNLKDGTCQEQKAGCKDYSTONKCSAQSGTSLCIVDS 1502
Qy 684 ---CG-----GAACTT-----GTCACTCT--- 698

1503 SCYTLGEKNCSITGTGLDHAQCCQSYSGAGTSITDGTCKQDLKSSCEQYAGTTASCTKTA 1562
699 --CAAACGA-----CRAAAGGAGCTT-----TTGC----- 721
1563 TSKYLOGSTCIITISNVATDCAKITGGAGTVTVDIQSYNTGGSANRARSACVQOQAQCS 1622
722 -----CACTGACTGGCCGCTGCTGACACCTC-----CAGAACGCA-- 758
1623 GYTVLTSCYKSGAGLCIASNTNTDTACVAASSATTCDAVLTGTYNANANVEKAGCTNN 1682
759 GGTGTC-----TGGCGC-----CCGTTCTGC----- 778
1683 GTTACVAKSCANAAGITFNHANCNSYLTCTVNSGGTACQTMIPKCADOTSTACLSVEG 1742
779 -CTGGGAGC-----CGGGAACCT----- 796
1743 ECVVIGTSCVRKTCDAIDATEDDDSECTVQOQSVCTVARLGACQARAACGTYKSLQCK 1802
797 -----C-----TCCTGCCG-----GAA-GCC 811
1803 FNTSGGKCFWNPINKICVDLNCNISTSTSFTHNECAVDSTLACTVRATNGAAVQGM 1862
812 G-GAC-----GGC----- 818
1863 ARGACASYTIEOCKTNASNGICVMNTNANLPAPACQDKSCTSPATSTTTHNDCYAYNT 1922
819 -----AGGATGGGCCCAACTT-----CG-----CCCTGC- 844
1923 ATVKCTVATPSNGNPTLGGCQOQTHACSSYDKEQCINANGDPGWNGTQCADRACS 1982
845 -----CCACTTGACT-----TC-----ACCAAT 863
1983 TAPATTDYDDTKRAVLSNKTIVSDIGECVEIPATCETMTQKCYSNKAGDPFYWTGT 2042
864 CCTTCTCTGAGA-CTAAACCTGTGCT-----C-----AGGAGCGAAGGACTGT 907
2043 TCITKSCDNAPDATATADECNTYLAGCTLDNVKCKTKVEDPAPATDALCKQALSTCTTN 2102
908 GAACTT-CT-----AGCCTGAAGAGC-----CAGAGC-TAGCTCT 940
2103 GTNCVTRGTCTFOALSQAGCVSSTTGOQCEWIPAVLNASNVVTPPAYCTIKNCSTAPITLT 2162
941 GGCCACAG---CT-----GGCGACGTCACTCTCTCCACCCACC-----C 981
2163 SE--AACAGYFTNCTTKNGGCVTKSTCAAVIDVACTTALNGTICAWDSQNKCRDKDC 2220
982 CCAAGTT-----CTA-AGGTC----- 996
2221 QDPSGTHAACQAOBRAGCTAGASCKCARVQNCETQVRSACIEGTNGPCLWIDKYVNTDG 2280
997 -----TTTTACAG--CGTGA-----GGTGGAGGAGGTGC- 1028
2281 SKGACFRYTSWKSILAMNSDTSCKWISNKCTTNGSCVGTLCSETNTDGGCVTYDGACI 2340
1029 -----TGCTCTCCAAACT---ATGCCAAGGC-----GG 1053
2341 QSVFALNSSDPKCPYISCADAFYTHSDQIASSKCTTNGTGTGCTCTCTGGAGAAGGTTTC 2400
1054 C-----GGCA-----GAGCTGGTCTTGGTCTCTCTCTGGAGAAGGTTTC 1092
2401 CYFNDKGAQYSSGAIVSTGVCTWDTISSCRDQSCADLTGTTHATCSQSLSCTSDGTTC 2460
1093 TGTTCCTCTGAT---TTATGAACCTC-----TA-----TAATAGATATAGG 1132
2461 L-LKGVCYSYTTQIACITAVGSDGSCYWELASATNNNTAKRLLTCAQINGTATNVCSV 2519
1133 TTTT-----GTAC-----CTT-TTTTAC-----AGGAGGCTACT--TTCT 1165
2520 ALSTCVSNGTACIPKANGSTVTSKACNSGGDLGICVFTQSTATGAAAGTCTCALMTACT 2579
1166 -----GTAAACA-----TCCGATGTATATTAACCTTTTATAAA 1199
2580 IANNDQTACAARDRCAWTAASGTGTAVASKCATHTCATNQA 2622

Best Local Similarity 32.9%; Pred. No. 5,1e-47;
Matches 397; Conservative 13; Mismatches 443; Indels 353; Gaps 55;

QY 17 AGGAAAGCAAACTCTGGAAGTAATGTTTTCAGTGAATTTTAAATCTTCTGCGCG 76
DB 84 AGAYASAAAABEQIVLVINAPQALLGRPLDGNATTPGAGAGGLIFGNGAGA 143
QY 77 AGAGCCCGCTCTCCCGGTATCAGCGCTTCTCATTTCTTTGAATCCCGCT--CCGC 134
DB 144 AGAPQAGG-----PGPAGLWNG-----GPGAGAGGGGTGAGGA 181
QY 135 GGTCTTCGCGCTCAGACCGGAGGAGCGCTTTTGCAATTAAGCGGCTGTGAACG 194
DB 182 GNLFGVGGAGVGA--GGTGGAGPGSLIW-----GGAGGVGAGGCTGGAG 232
QY 195 CC--CAGGCGCGCGGGGGCGG-----GGCCAGGCG--GGCCATTTTGAATAAGAGCG 246
DB 233 RAELLFGAGGAGGAGTGGPGATGCTGHHGVDGGWLPAGGAGGAGCGGAGGAG 292
QY 247 TGCCTTCCAGGAGGAGCTCTAATAGTGACCGCGCGGGA--CGTGGCGGCTTCAGG 303
DB 293 SD-----GAGLGG-----TGTGTGGAGGAGGALLGAG-----GGGLGAGG 334
QY 304 TCAGTGACGAGCTCTTTTGGTTTCTTCTTTGGGGACCTCTGAGCTCACTCC 363
DB 335 Q---CGMGAGGAGADNPTGIGG-----TGGDGGTGGGAGGAGG----- 372
QY 364 CAG---CATGAGGCGCTGAGCCCGGTGCGGCTGTACGAGCGGTGT-----CCTGC 415
DB 373 AGQLFSAGAGNAGVGA---GGG-----GAGGAGGAGADADQPGATG- 415
QY 416 CTGTCGGAACGAGTCTTGCCCATCGCCCGGGCGCGAGG-----GAAGGCGCGGAGCTGA 471
DB 416 GTGFAGGA--GGAG-----GAGGSSGAGGTGSGGAG--TGCGGAGGA 456
QY 472 GGAGCGCTGAGCT--TGCTGGAGGACATGAACCACTGCTACTCCCGCTGCGGAACTG 529
DB 457 GGAGADNPTGIGTGGDGTGAAGA-----GAGGAGGAGT 492
QY 530 GTACCGGAGTCCGAGAGGAGCTACGCTAGCCAGTGAATCTTACGCGGCTATC 589
DB 493 GTGGMIGTGNAGV--GAGG-----QGDDGAG-----GAGADADQ 527
QY 590 GACTACATCTCGAGCTGAGTCTGCGCGGAGGAGCCGCTGAGCCCGCTGATGC 649
DB 528 PGATG-----GTGFAGGAGGAGGSSGAGGTG-----SGGA-----GGTGG- 566
QY 650 CCCACCTTCCATCCAGACAGCGGAGC-----TCGCTCCGGAACCTTGTCACTCCA 701
DB 567 -----QGAGGAGGAGADNPTGIGTGGDGTGGAAGAG----- 601
QY 702 ACAGAAAAGAGCTTTTGCACCTGCTCGCGCTGTCTGACACCTCCAGAAACGAGT 761
DB 602 -AGGAAGTGTGMLGTTGNAGVGGAGGCGDGGAG-----GAGG-----AGGSSG-AGT 650
QY 762 GCTGCGCGCTTCTGCTCGGACCCCGGAACTCTCTGCGGAGAGCG--GACGGCAG 820
DB 651 NGSGAG--GTGGCGGAGGAG-----GAGADNPTGIGTGGDGTGGAAGAGGAGGAG 702
QY 821 GATGGGCCCCAACTTGGCCCTGCCACTTGACTTCAACAAATCCCTTCTGAGAGCTAA 880
DB 703 TGTGG-----MIGTGNAGVGGAGGCGGAG-----AGGAGADADQ 739
QY 881 ACTTGT-----GCTCAGAGGAGGAGTGAATCTGAGCTTGAAGCCAGAGCTAG 936
DB 740 PGATGTFAGGAGGAGGAGGAGGAGTGGAGTGGG-----GAGGAGTGFSGNSNG 796
QY 937 CTCTGCCACGAGCTGGGCGAGCTCACCTGCTCCCAACCCCAACCTTCTAAGTTC 996
DB 797 TGGTGG-----VGTGGDGNAGT-----GAGDPGKGGT 826
QY 997 TTTTCAGAGCTGGAG--GTGTGAGGAGGAGTGGCTGCTCTCCAACTATGCCAGGCGCG 1055

DB 827 GTGSGGAG--GSGANFNNGTGTGTGTGTGGKGG-----MGGIAGDGGPGGDG 874
QY 1056 GCAGAGCTGTCTCTCTGCTCTCTCGAGAGAGGTTCTCTTCCCTGAT--TTATGACT 1113
DB 875 GNAGVGKGG--JNNGSGGTGTGTGPGSGGAPGSGTGKGGAGGCGDGDGAAAT 932
QY 1114 CTATATAGAGTATAGGTTTGTACCTTTTACAGGAGGAGTCTTCTGTAACTAT 1173
DB 933 GVDGSGDGGNG-----GNGGNGGTGTG-----SPGGLGAGG-----TGLGAGGA 973
QY 1174 GCGATG 1179
DB 974 GGGADG 979

RESULT 13
076973
ID 076973 PRELIMINARY; PRT; 2721 AA.
AC 076973;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Surface antigen G.
GN 51G.
OS Paramesium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramesium.
NCBI_TaxID=5888;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=51;
RX MEDLINE=99038212; PubMed=9819394;
RA Duhaucourt S., Keller A.M., Meyer E.;
RT "Homology-dependent maternal inhibition of developmental excision of
Internal Eliminated Sequences in Paramesium tetraurelia.";
RL Mol. Cell. Biol. 18:7075-7085(1998).
DR EMBL: AJ010441; CAA09180.1; -
DR GO: GO:000524; P:ATP binding; IEA.
DR GO: GO:0004572; P:protein kinase activity; IEA.
DR GO: GO:0007275; P:development; IEA.
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR002895; Paramesium SA.
DR InterPro: IPR003659; Plexin-like.
DR InterPro: IPR000719; Prot kinase.
DR InterPro: IPR001680; WD40.
DR Pfam: PF01508; Paramesium_SA; 34.
DR SMART: SM00433; PSN; 32.
DR PROSITE: PS00107; PROTEIN KINASE_ATP; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 1.
SQ SEQUENCE 2721 AA; 280343 MW; 630780EDB7B76242 CRC64;

Query Match 11.9%; Score 894.5; DB 5; Length 2721;
Best Local Similarity 21.6%; Pred. No. 1.3e-44;
Matches 528; Conservative 51; Mismatches 607; Indels 1257; Gaps 127;

QY 3 TCTGGG-----GTGC-----TGCCAGGAAAAAGCA--AATTCGGAAGTT 40
DB 191 TKTGGGCVDSNNCSQTLLEIQCVNKLKTTACYWDGACKDKRICDNAPTSLTDDACKT 250
QY 41 AATGGTGTTCAGTGATTTTAAAT-----CCTTG----- 69
DB 251 FRDGTCTKANGCVTRTTCAAAATQASCKNSGGDCYWTGTACVDKTCANAPVTMT 310
QY 70 ---CTG---CGGAGAGGCC--GCC-----TC--TCC----- 92
DB 311 NSACAGFTVGTCTKSGGCVANGACSVANVOACVKNSSNFDICWDTTCKEKTANAPT 370
QY 93 -----CCGTATC-----AGGCTTCTCTCACTTTTGAA-----GCCG----- 129
DB 371 NNTDLCTSLVSLCTVKSGGCGNRTCANAPITLTNDACEAYLTGNNCITKTGGCVTN 430
QY 130 -TCC-----CGGTCTTCGCGCTCAGACCAGC-----CGGAGG 161

QY 218 ---CGA---GGCGG-----GC-----CATTTTG----- 234
Db 605 AKTTATTSDGTYKSGCVANNPNVNGSIQGGQDLFTTCAARKSTENCEIIRTGPFPTCLMN 664
QY 235 RATAAGAGGCGTGCTT-----CCAGGACGGCTCTATAAGTAGACCCGCGG 282
Db 665 SATSACVKEKSCSTASVTITTTGFLTVFSNTNCLAYLSNSACIANNTADSCIPKPSGCSLT 724
QY 283 CGAGGGTGC-----GGCGGTGCGAGTC-----ACTGTAGCGGACTCTTTTGGTTTCTTTC 335
Db 725 TSANGVSGSKANGDCVWNGTTCTVDRTCSNISLSHSGNGVLTNC--TVNNAKICOSLA 782
QY 336 TCTTTGGGGCAC-----CTCTGCAC--TC-----ACTC-----CCCA 365
Db 783 TACTSYGSENCKQTSAGKTCIWTGACRNATCLDASDNTSPDSGECSAVPTPSDTCTV 842
QY 366 GCATGAAGCGCGTGAGC-----CCGCTGCGCGGCTGCTACG-----AGCGG--- 406
Db 843 LYKTGAV-GGVTRSANCSVDYVQAQCVRLTAASDDCTWKSCHKFSNTYLSGACSTFQ 901
QY 407 GTGTGC-----TGC-----GTGTGC--GAA--CGCAG--TCT-----GGCCAT 438
Db 902 GTKTLCEAIKGTCTNAGAVETDCTFFCAVVTGSLTHAGCQGYSTTCTANSGGTACLT 961
QY 439 CGCCC-----GGGC-----CGAGGGAAGGCG-----CGCGCAGCTGAGAGCCGCT 480
Db 962 LAACSTYTAAGDCVVRTDGLSCFVNAGASPRACQDITSTNCNALASLTGATHASCOAYS 1021
QY 481 GAGCTTGCTGGAC-----GACATGAACAC-----TGCTACTCCCG--CTGCGG 523
Db 1022 TA-CTISDGTACQAIACSSYSALTACTKRTDGLKCFWNTGATPAACQDITSTNCNA 1080
QY 524 GAACTGTACCGGAGTCCGAGAG-GCACTAGCTT-----AGC-----CAGGTGA 570
Db 1081 LASLTGATHATQAYSTACTSISDGTACQAIQACSSYSALTACTORTDGLCFWNTGAT 1140
QY 571 AATCTTACAGCGG-----TCATCGACTACATCT-----CGACCTCAGGTAGTCTCT 618
Db 1141 PAAACQDITSANCNALASLTGATHASCOAYSTACTSISDGTACQALQACSSYSATCT 1200
QY 619 GGCGGAGC-----CAGCCCT-----GGAC--- 638
Db 1201 QRTDGLKCFWNTAATPAACQDITSANCSLTGLSSATHATQAYSTACYNLAGNACQA 1260
QY 639 ---CCCCTGATGCGCCCACTTCC-----CAT---CCAG-----ACAGCC--- 673
Db 1261 FSTCEALTGSNLTWTICQAFSTTCSVKRDGTGCVTIQSACTGYTTIANCVRSTAGYCTAN 1320
QY 674 -----GAGTC-----GCTCCGGAACCTTGTCATCT----- 698
Db 1321 SGDSACQAIASSTTCEQIKLGSSPAFDDTKNTFKTGCIALGHSQCQTKTCAKNTPPAH 1380
QY 699 ---CCAACGACA--AAAGAGCTTTTGCACTGACTC-----GGCGGTGCTCT----- 741
Db 1381 SDONTWLTCTSNVSSPTACTTWAATCASLTTSTCVYAVEGCVWSGTSCVAKTCDTAS 1440
QY 742 -----GACAC--CTCCGAAAGCAGGTGC-----TGG----- 766
Db 1441 ADTSFDSHAECIAYLSTCTVARTGCGQARATCASYSQCKFNSTGKCFWNPNNKTCV 1500
QY 767 ---CGCCCGT-----CTGCTCGGACCCCGGGAACCTCT-----CCTGC 803
Db 1501 DLNCGNIEATTTDTHAECAVDFTQLLCTVTRATNGAIVPGCMARGACSSYSIEDQCTNP 1560
QY 804 CGGAA-----GCGGAGCGGAGGATG 825
Db 1561 SGGVCMVNTLTTPVQDKSCTTAPTATATHADCDYSFSTATIKCTVVAFTDNGGAHL 1620
QY 826 GGCCCCCACTT-----CG-----CCCTGC-----CCACTTG 851
Db 1621 GGCQQTACSSYTHOEQCFRNATGDLGNGTQCADKSCATAPATTTDNDKCRAYFNN 1680

QY 852 ACT-----TC-----ACCAATCCCT-TCCTGGAGACTAAAC 882
Db 1681 KCTVASSGGCGVDIPPTCESMTQKQCVSDKTRGSCVWNGTACITRTCNAPDSTASAEDC 1740
QY 883 CTGGTGCT-----C-----AGGAGGAGGAGTGTGACTTT-GT-----AGCC 919
Db 1741 NTYLAGCTLDSVKCKTKVCEDEFATFADLCKSALSTCTTNGTNCVTRGTCFQANQAGCV 1800
QY 920 TGAAGAGC-----CA-----GAGCTAGCTC 939
Db 1801 TSSANVQCEWMPAVGYNQAYCTVKTCTNATPATLTSASASYFTNCTTKNGGCGCVTKSTC 1860
QY 940 T-----GGCC-----ACC-----AGCTGGCGGAGG- 959
Db 1861 SAVTIDVACTTALNGTVCAWDSANQKCRDKDQDFSGTTHAACQOTRAGCTAGANKCAR 1920
QY 960 --TC-----ACCTGCTCCC-----ACCCACACCCCA--AGTTCTAAGGT 995
Db 1921 VQNCQETIRSAIEGTNGFCLMNNVNSDGSKGACFRVTSCKSLAWSNDSKCKWISNQ 1980
QY 996 CTTTTCAGAGC-GTGGAGGTGTGGAAGAGTGGC-----TGCT----- 1032
Db 1981 CTTN---GSCIGITLCSNETDGGCVSYDGGACIQSVPALNSDPKVKXPYSCADAFY 2037
QY 1033 ---CTCCAACT--ATGCCAAGCGGC---GGC---AGAGCTGGTCTTCTGTGCTC 1077
Db 2038 TTHSDCOTASKKCTTNGTTGCIAGLACSSYTTQAGCYFNDKGLLTSGLTST--GICTW 2095
QY 1078 CTTGAGAGAGGTCTCTTGCCTG--ATTATGAAC---TCTA---TAATAGAGT-- 1125
Db 2096 DTTASSCRDQSCADLTGTTHATCSSQLSTCTSDTSLVKGACTSYTTQACTTAVGSDG 2155
QY 1126 -----ATAT-----AGGTTT-----TGTAAC---CTT-TTT 1146
Db 2156 ICYWEASSSTNNNTAKRLLACADIQNGTSTNVCAVALSCSVNGTACIANKSTYTK 2215
QY 1147 TAC-----AGGAGGTG-----ACTTTCTGTACAA-----TCGGATGT 1180
Db 2216 TACNSGGLDGICVFTQSTATGAAAGTGT'CALMTACTTANSQVACQAAKDRCSMTAASGT 2275
QY 1181 ATATTAACCTTTTATAAA 1199
Db 2276 GTTAVASKATHTCATNQA 2294

RESULT 12
QYTWB8 PRELIMINARY; PRT; 992 AA.
AC QYTWB8;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE PE-PGRS family protein.
DE PE_PGRS57 OR M83543.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; EX248346; CAD95729.1; --
KW Complete proteome.
SQ SEQUENCE 992 AA; 77324 MW; D0140B51C0042121 CRC64;
12.3%; Score 929.5; DB 16; Length 992;
Query Match

[illegible]

```

Db      1962 GYGYSNSAAAAAAGGAGGGYGGYGGWGDGYSDSAAAAAAAAAAAGSGAG 2021
Qy      1199 AAG 1201
Db      2022 GAG 2024

RESULT 11
Q27167 PRELIMINARY; PRT; 2395 AA.
ID Q27167
AC Q27167;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE 51B type surface protein.
OS Paramacium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramacium.
OX NCBI_TaxID=5888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STOCK 51;
MEDLINE=95098630; MedMed=7800503;
RA Scott J., Leeck C., Forney J.;
RT "Analysis of the micronuclear B type surface protein gene in
RT Paramacium tetraurelia.";
RX Nucleic Acids Res. 22:5079-5084(1994).
RL [2]
RR SEQUENCE FROM N.A.
RS STRAIN=STOCK 51;
SCOTT J.M., Leeck C.L., Forney J.D.;
RA Genetics 133:189-198(1993).
RL EMBL; U07603; AAA81947.1; -
DR EMBL; L04795; AAA16710.1; -.
DR FIR; S50820; S50820.
DR GO; GO:0007275; P:development; IEA..
DR InterPro; IPR002895; Paramacium SA.
DR InterPro; IPR003659; Plexin-like.
DR InterPro; IPR001680; WD40.
DR Pfam; PF01508; Paramacium_SA; 31.
DR SMART; SW00439; PSA; 28.
DR SMART; SW00423; PSI; 3.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR SEQUENCE 2395 AA; 246430 MW; 8632E022649BD61 CRC64;

Query Match          12.3%; Score 930; DB 5; Length 2395;
Best Local Similarity 23.4%; Pred. No. 9,8e-47;
Matches 481; Conservative 67; Mismatches 637; Indels 874; Gaps 111;

Qy      3 TCTGGGTG-CGCCAGGAAGAAGCAAT-----TCTGGAAGTAATGGTTTGAGT 53
Db      248 TCXSFRTDGFCTTKANGCGCVTTTCARATIQAACVKNSSNDCVWTGTACVDKTCINA-- 305

Qy      54 GATTTTAAATCCTTCTCT---GGCGGAGA-----GCCCGCCCTC---TC 91
Db      306 PATMTNSACAGFVTCTIKSGGGGVSNAGCSAANVOAACVKNTSGTDCIWDTTCKEKC 365

Qy      92 C-----CCGATATC-----AGCGTCTCTCATCTTTGA-----ATCC---GC 126
Db      366 ANAPTNNTHDLCTSYLSCTVKSGGCGOPTCANAPPTITNDACEAYLPANNCTKSG 425

Qy      127 GGC-----TCC-----CCGCTCTTCGGCGTCAACACGCGGAGGAAGCCGTGTTT 171
Db      426 GGCVENTTCAVITLEAACVKNSSGATCFWDSASSTCKDKTCAN-APSTNTNDLCVAFLS 484

Qy      172 GCRAATTATGAC-----GGGC-----TGTAAC-- 193
Db      485 SCTVNSTINACVDKTCENSLAQICDKLNKACIKWGKCYKRECVLASFTTAHADCQT 544

Qy      194 ---GCCCA-GGGC-----CGG-----CG-GGGCGGGGC-- 217
Db      545 YDVGCTLSNTGTGCVPLFKCAITAEACNIRLQVTSVGKVSYPLOCWNGSSCIDKACST 604

```



```

QY 146 TCAGCAGCGCGGAGGAGCCTCTTTGCAATTTAAGC-----GGGCTGTGACGCC 197
Db 524 AAAAAAAAAAGSAGCGGCGGCGSSAAAAAARAGHDSNAGSAAAAA 583
QY 198 AGGCGCGCGCGCGCGCGGCG-----CGAGCGCGG-----C 227
Db 584 AASG-AGSGGGYGWGDGYSAAAAAAGSAGGAGGAGGCGGCGGSDA 642
QY 228 CATTTTCAATAAGAGCGCTGCTTCAGCAGCGCTCTATAGTACCGCCGCGGAGC 287
Db 643 AAAAAAAAAAASGARGSGG-----YGGYSDSAAAAAAGSAGGAGG- 694
QY 288 GTGCGCGCTGTGAGGCTACTGTAGCG-----GACTTCT----- 321
Db 695 GWGDG-YGSDSAAAAAAGSAGGAGGCGGCGGSDSAAAAAASARRAGHD 753
QY 322 TTTGGTTTTCTTCTTTGGGCACCTCTGAGTCACTCCCGAGCATGAGCCCTGAG 381
Db 754 SAAGSAAAAAAGSAGGSGGGYGWGDGYSAAAAAAGSAGGAG 813
QY 382 CCGCGTCCGCGG-----CTGCTACGAGCGGTGTGCTGCTGCGGAGTC 429
Db 814 ---GGYGWGDGYSAAAAAAGSAGGAGGCGGCGGSDSAAAAAAGSAGG 870
QY 430 TCTGCCATCGCCGCGCGGAGGAGCGCGCGAGCTGAGGAGCGCTGAGTCTGCT 489
Db 871 SGAGVGCGYGWGDGYSAAAAAAGSAGGAGGCGGCGG-----S 917
QY 490 GGACGACATGAACCA-----CTGCTACTCCGCGCTCGGGAACGTGATCCGAGTC 541
Db 918 GSSAAAAAAGSARRAGHDSNAGSAAAAAAGSAGGSGGGYGWGDGYS 977
QY 542 CCGAGAGGCACTGAGCTTACCGAG---GTGG-----AAATCTACAGC-GGTCAT 588
Db 978 SAAAAAAGSAGGAGGCGGCGGSDSAAAAAAGSAGGAGGSGGGY 1037
QY 589 CGACTACATCTCGACCTGC-AGTGTAGTCTG---GCCGAGCCAGCCCTGACCCCT 643
Db 1038 YGSDSAAAAAAGSAGGAGGCGGCGGSDSAAAAAAGSAGGAGG---S 1091
QY 544 GATG-----GCCCGCCCTTC-----CCATCCAGCAGCGGAGCTGCG 681
Db 1092 GAGRGCGGYSAAAAAARAGHDSNAGSAAAAAAGSAGGAGGSGGG 1151
QY 582 TCGGAACTTGTCTATCTTCAACGACAAAGGAG-----CTTTGCCACTGAC 728
Db 1152 YGWGDGYSAAAAAAGSAGGAGGCGGCGGSDSAAAAAAGSAGG 1211
QY 729 TCGGCGCTGTCTG---ACACCTCCAGACGCGAGGTGCTGCGCGCTTCTGCTGGAC 785
Db 1212 GAGGCGGCGGSDSAAAAAAGSAGGAGGCGGCGGSDSAAAAAAGSAGG 1270
QY 786 CCGGGAACCTCTCTGCGG-----AAGCCGCGGCGGAGGAGG-GCC 829
Db 1271 AAGSGAG-----GAGGCGGCGGSDSAAAAAAGSAGGAGGCGGCGG 1322
QY 830 CCACTTTCGCGCTGCGGCACTTCACTTCACTTCTCTGAG----- 875
Db 1323 SAAAAAARAGHDSNAGSAAAAAAGSAGGAGGCGGCGGSDS 1382
QY 876 -----ACTAACTCTGTTG-----CTCAGGAGCGAGG----- 902
Db 1383 AAAAAAAGSAGGSGGGYGSAAAAAAGSAGGAGGCGGCGGSDS 1442
QY 903 -ACTGTCACTGTAGCTCAAGAGCGAGAGCTAGCTTGGCCACCGAGCTGGCGGCGCTC 961
Db 1443 YGSDSAAAAAAGSAGGSGGGYGSAAAAAAGSAGGAGGAGG--- 1500
QY 962 ACCCTGTCCACCCCGGAGTCTTAAG-----TCTTTTCAGA 1004
Db 1501 -----GYSDSAAAAAAGSAGGAGGCGGCGGSDSAAAAAAGSAGG 1555

```

```

QY 1005 GCCTGAGGAGTGGA-----AGGAGTG 1026
Db 1556 GAGRGCGGCGGSSAAAAAARAGHDSNAGSAAAAAAGSAGGAGG 1615
QY 1027 ---GCTGCTCTCAAACTATGCAAG-GCGGCGGAGAGCTGCTTCTTCTGCTCTCTGG 1082
Db 1616 YGWGDGYSAAAAAAGSAGGAGGAGGCGGSDSAAAAAAGSAGG 1675
QY 1083 AGAAA-GGTTCTGCTGCGCTCTTATGAACTC-----TATAATAGAGT 1125
Db 1676 SGAGRGCGGCGGSDSAAAAAAGSAGGAGGCGGCGGSDSAAAAAAGSAGG 1735
QY 1126 ATATAGCTTTTGTACCTTTTACAGGAAG-----GTGACTTCTGTAACT 1173
Db 1736 AARRAGHDSNAGSAAAAAAGSAGGCGGSDSAAAAAAGSAGG 1795
QY 1174 GCGAT-----GTATATTAACTTTTATAAAGT 1202
Db 1796 GAGSGCGGCGGSDSAAAAAAGSAGG 1833

RESULT 10
Q964F4
ID Q964F4 PRELIMINARY; PRT; 2655 AA.
AC Q964F4;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Fibroin.
OS Antheraea yamamai (Japanese oak silkworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Saturniidae; Saturniinae; Saturniini; Antheraea.
OX NCBI_TaxID=7121;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSDUCED=mariner-like element;
RA Lee J.-S., Kang S.-W., Suh D.-S., Jin B.-R., Kim Y.-S.,
RA Lee K.-S., Goo T.-W., Yun E.-Y., Kim K.-Y.;
RT "Complete nucleotide sequence of fibroin gene cloned from Antheraea
RT yamamai."
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF255500; AK83145.1;
SQ SEQUENCE 2655 AA; 218672 MW; 89B8810EF9CC4CA6 CRC64;

Query Match 12.9%; Score 970.5; DB 5; Length 2655;
Best Local Similarity 25.0%; Pred. No. 4.6e-49;
Matches 375; Conservative 54; Mismatches 721; Indels 353; Gaps 39;

QY 1 GATCTGG---GTGCTGCCAGGAAAGCAATCTGG-----AGTTAATGG 45
Db 573 GAGSGCGGCGGSDSAAAAAAGSAGGAGGCGGSDSAAAAAAGSAGG 632
QY 46 TTTTG---AGTGATTTTAAATCTTGTGCGGAGAGCGCCCTCTCCCGGTATCA 101
Db 633 AARRAGHDSNAGSAAAAAAGSAGGCGGCGGSDSAAAAAAGSAGG 690
QY 102 GCGC-----TTCTCTATCTTTGAATCCGCGCTCCCGCTCTTCGC 144
Db 691 GSAGAGGCGYGWGDGYSAAAAAAGSAGG---GAGGCGGSDSAAAA 747
QY 145 GTCAGACCGCGGAGGAGCGCTGTTGCAATTTAAGCGGCGTGTGAAGCGCGGCGC 204
Db 748 AAAAAAAGSAGGAGGCGGSDSAAAAAAGSAGGAGHDSNAGSAAAA 807
QY 205 GCGGCGCGGCGGCGGCGGCGCTTTTGAATTAAGAGGCGGCTCTCCAGGAGGCTC 264
Db 808 AASGAGSGCGGCGGSDSAAAAAAGSAGGAGGAG---GDTWGDGCGG 863
QY 265 TATAGTACCGCGCGGAGCGCTGCGCGCTTCAAGTCACTGAGC----- 314
Db 864 DSAAGSAGGAGGCGGCGGSDSAAAAAAGSAGGAGGCGGCGG 922

```


OS	Mycobacterium tuberculosis.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CDC 1551 / Oshkosh;	
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,	
RA	Peterson J., DeBoy R., Dodson R., Gwim M.L., Haft D., Hickey E.,	
RA	Kolony J.F., Nelson W.C., Weidman L.A., Ermolaeva M.D., Salzberg S.L.,	
RA	Delcher A., Utterback T., Waidman J., Khouri H., Gill J., Mikula A.,	
RA	Bishai W.;	
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and	
RT	laboratory strains.";	
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AB007163; AAK47971.1; .	
DR	TIGR; MT3612; .	
DR	InterPro; IPR000084; PE region.	
DR	InterPro; IPR002173; PfkB.	
DR	Pfam; PF00934; PE; 1.	
DR	ProDom; PD001223; PE region; 1.	
DR	PROSITE; PS00583; PFKB KINASES 1; 4.	
SQ	SEQUENCE 1715 AA; 133832 MW; E857B2A774C771B0 CRC64;	
	Query Match 13.0%; Score 979; DB 16; Length 1715;	
	Best Local Similarity 31.1%; Pred. No. 1e-49;	
	Matches 419; Conservative 21; Mismatches 527; Indels 382; Gaps 58;	
QY	1 GATCTGGGTCGTCGAGAGAGCCCGCTCTCCCGGTATCAGCGCTTCTCATTTCT	56
DB	221 GGLWGGGAGGCGGAGGTGGAGRAELLFGAGGAGGATGGPGATGGTGGHGGVD	280
QY	57 TTTTAAATCTCTCTGCGGAGAGCCCGCTCTCCCGGTATCAGCGCTTCTCATTTCT	116
DB	281 GGLWGGGAGGCGGAGGAGSDG-----GALGGTGGTGG-----	316
QY	117 TTGAATCCGCGCTCGCGCTCTTCGCGTCAGACCGCGGAGGAGCGCTTTTGCAAT	176
DB	317 -TGAGAGGRRGALLIGAGGQ-----GGLGAGGQ-----GGTGGAGG--DGVLGVEG	363
QY	177 TTAAGCGG--GCTGTGAACCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	233
DB	364 GKGGVGVVAGLGGAGGAGOLFAGAGAGVGVGGTGGGAGGAGAGADAPASTGLT	423
QY	234 GAATAAGAGCGCTGCTCCAGG--CAGGCTCTAATAGTACCGCGCGCGCGCGCGCG	291
DB	424 GGTGFAGGAGGVGG-----OGGNATAGGINGSAGGAGTGGGAGGAGGAGGAGN	474
QY	292 GCGCGTTGCAAGTCACTGTAGCGGACTTCTTTTGGTTTCTTCTTTTGGGCGACCT	351
DB	475 --GIGADGGAGTGGNAGAGAGGAG--TGGTGGVVGAGAGAGIGTGGGQGG--	524
QY	352 GGACTACTCCGAGCATGAGCGCTGAGCCCGGTG--CGCGGTGC-----TACGA	402
DB	525 GGAGSAGTD-----ATATGATGTFSGGAGGAGGAGGNTGVGGTNGSGGCGGTG	579
QY	403 GCGGCT-----GTGCTGCTGTGCGAACCCAGTCTGGCCATCGCCGCGCGCGAG	453
DB	580 GGAGGVGADNPTGIGTGGTGGGAGGAGGAGADATATGATGTT--GFAGGAGGAGG	638
QY	454 GA-AGGCGCGGAGTGGAGCGCGCTGAGCTGCTGGAGCAGCATGAAACCATGCT	512
DB	639 SSGAGGTNGSGAGGTGGGAGGAGGAGADNPTG--TGGAGTGGTGGAGGAG--	689
QY	513 CCGCGCTCGGGAAGTGTACCGCGAGTCCGAGAGGCACTCAGCTTACCGAGTGGAAA	572
DB	690 -----GGAGGAGTGGTGGAGVSVG-----NAGIGG-----TGG--TGGV	727
QY	573 TCCTACAGCGCTCATGCACTACATTTCTGACCTGAGGTAGTCTCGCGCGAGCCAG	632
DB	728 -----AGAAAGSSATGAGFA-----GGAGGAGGAGGNSGVGGTNGSGAGG	774
QY	633 CTGAGACCCCTGATGGCCCCCAGCTTCCCATCCAGA-----CAGCCGAGCTCG	685

DB	775 -KGG-----TGGAGG-----: : : : : :SGADNPTGAGFAG--GAGGTGGAAGAG	808
QY	686 GAACCTTCTATCTCCACAGCAAAAAGGA-----: : : : :GCT	716
DB	809 GAGGATGTGTGGVVGATG--SAGIGGAGRGDGGDASGLGLSLGSPDGGGQGGAG	866
QY	717 TTTCCTACTGACTCGGCGGTCTCTGACACTCCAGAACCGAGGTGCTGCGCCGCTTCT	776
DB	867 GSAGAGGAGGAGGAGGAGG--ATGA-----AGLDNGGVGGDGGAG--GAAGN	912
QY	777 GCCTGGGACCCCGGGAACCTCTCTGCGCGGAAGCCG-----GACGCGAGGAGTGGCC	829
DB	913 GGNAGVLTAKAGDGA-----AGNGGNGGAGGAGGAGDNNFNGGQGGAGGQGGG	964
QY	830 CCAACTTGGCCCTGCCACTTGACTTCACCAATCCCTTCTCTGGA-----GA	876
DB	965 -----LGGASTT-----SINAGGAGGNGGTGKGAGGA	994
QY	877 CTAAACCTGTGTG-----CTCAGGAGCGAAGG-----ACTGTGAAC	912
DB	995 GTLVGGSGGTGGDGDAGAGGGGGFGAAAGKAGGGNGGVGGDGGEGASGLGLSLG	1054
QY	913 -----TGTAGCCTGAAGAGCCAGAGTAGCTCTGGCCACCAGCTG-----GGCAGCTCA	962
DB	1055 GGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1114
QY	963 CCCTGCTCCCAACCCCAAGTCTTAAGGTCTTTTACAGAGCT---GAGGTGTGGA	1019
DB	1115 -----GFGAGVPGDGGDGN---AGFGAGVPGDGGIGTGTGAGG	1150
QY	1020 AGGAGT-----GGCTGCTCTCCAACTATCCCAAGCGCGG-----	1056
DB	1151 AGGAGADPDFIDGGGAGGHHGGGKGKGLNSTGLASAGSDG--GNGGAGGAGNGGD	1208
QY	1057 ----CAGAGCTGGTCTTC-----TGGTCTCTTGGAGAAAGTCTCTGTTCCCTG	1105
DB	1209 GDGFIGSGGTGGTGGAGVGLANTGTAGNAGIGAG--GRGDDGAGDSGALSQDNG	1267
QY	1106 TATGAATCTATAATAGATATAGTTTGTACCTTTTACCTTTTACAG-----GAA	1154
DB	1268 FAGGCGGQ--GGVGNAGAGGAGG--AGGTGCTGGAGGAGPATLIGPDGDDGGGGG	1325
QY	1155 GGTGACTTCTCTGAACAATGCGATGTATA 1183	
DB	1326 GGNAGFAGVPGDGGIGGTG--GAGGAGGA 1353	
RESULT 8		
Q53557	PRELIMINARY; PRT; 1079 AA.	
AC	O53557;	
DT	01-JUN-1998 (TrEMBLrel. 06, Created)	
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	PGS-family protein (Fragment).	
GN	RV3512 OR MT023.19.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H37RV;	
RX	MEDLINE=98295987; PubMed=9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,	
RA	Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Davlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,	
RA	Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,	
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;	

QY 1147 TACAGAGAGTGTCTTCTGTAAACAATCGATGTATAT-----TAAACTTTTAA 1195
 Db 1584 GSGDGGAGGGAAGTGTGTGGD-GLTGTGTGGGGTGGGNGGNGGNGADNTANMTA 1642
 QY 1196 TAAAG 1201
 Db 1643 QAGSDG 1648

RESULT 6

Q7TWC3 PRELIMINARY; PRT; 1460 AA.
 AC Q7TWC3;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PE-PGRS family protein.
 GN PE-PGRS54 OR MB3538.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12789972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monseme C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.R., Gordon S.V., Hewison R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248346; CAD95724.1; -.
 KW Complete proteome.
 SQ SEQUENCE 1460 AA; 113187 MW; 180D6DB7DD4BDBA9 CRC64;

Query Match 13.2%; Score 998.5; DB 16; Length 1460;
 Best Local Similarity 31.0%; Fred. No. 6.5e-51;
 Matches 429; Conservative 9; Mismatches 576; Indels 371; Gaps 52;

QY 6 GGGGTGCTGCCAGAAAAGCAAAATCTGGAAGTAAATGTTTTC-----AGTGATTT 58
 Db 161 GNGGPGAGGGGGGTGGAGGAGGWLFGVGGAGVGGAGGTGGAGGPGGLINGGGGAGV 220
 QY 59 TTAATCTTCTGTCGCG-----AGAGCCCGCTCTCCCGGTATCAGCGCTCTCT 110
 Db 221 GGAG-----GGTGAGGABLLFGAGAGG-----AGGAGTDGFGATG--- 259
 QY 111 CATCTTTGAATCCCGCTCCGCGTCTTCGCGTTCAGACCCAGCCGAGAGAGCCTGT- 169
 Db 260 -----GTGCHGVGGDGGWLPFGAGAGAGGAGGAGSDGALGALGTGTG 305
 QY 170 -----TTGCAATTAAGCGGCTGTGAAGCCGAGG-CCCGCGGGGCGG-GGCCGA 220
 Db 306 GTGGAGGAGGAGLLGAGGGGGLGAGGGGTGGAGGDLGAGGGTGGAGGAGGAGGAGL 365
 QY 221 GCGGGCCATTGTAATAA-----AGAGCGGTGCTTCCAGGCGAGCTCTATAAGT 271
 Db 366 GGAGGAAGQLFSAGGAAGAVGVGTGGGAGGAGGAGGAGGAGGAGGAGGAGTGNAGA 425
 QY 272 GACCCCGCGGAGCGTCCGCGTTCGAGTCACTGTAGCGGACTCTTTTGTGTTT- 330
 Db 426 GGAGGAAGTGTGTVGAGGAGAGTGTGGGQGG-----AGGAGSAGTATATATGATGTFG 481
 QY 331 -----CTTTCTCTTTGGGCACTCTGCACTCACTCCCGAGCATGAAGGCGC---TG 379
 Db 482 GAGGAGGAGGNTGVGTNGSGGQGTGGAGG-----GGAGGAGVADNPTG 526
 QY 380 AGCCCGGTGCGGGCTGCT--ACAGGCGG--TGTCGTCTGTGTCGGAACGCACTCTG--- 433
 Db 527 TG-GAGGTG-GAGG-TGGTGAAGAGGAGGAGVGTGGTGGVGVGVNAGTGGTGGKAGAG 583
 QY 434 -GCCATCGCCCGGGCCGAGG-----GAAGGCGCCGCGAGCTGAGGAGCGCGCTGAGCT--T 486

Db 584 TGFAGGAGGAGGGSSGAGGTNGSGAGG---TGCGGAGGAGGAGADNPTGIGAGGT 640
 QY 487 GCTGAGCAGCATGAACAACCTGCTACTCCCGCTTCCCGGGAAC--TGCTACCCGAGTCCCG 544
 Db 641 GGTGGAAGAGGAGGA--IGTGGT-----GGAVGSVGNAGIGTGTGT---GAVG---G 684
 QY 545 AGAGGCACCTCAGCTTAGC-----CAGGTGAAAATCTCTACAGCG 582
 Db 685 AGGAGAAAAGSSATGAGFAGGAGGAGGAGNSVGGTNGSGAGGAGGAGGAGGAGGAGG 743
 QY 583 CGTCATCGATACATCTCTGACCTGAGCTAGTAGTCTTGGCCGAGCCAGCCCTG----- 635
 Db 744 SGADNPTGAGFAGGAGGTGAAGAGGAGGATGTGGTGGVVGATGSGAGIGAGGAGGAGG 803
 QY 636 ----GACCCCTCATGCGCCCGCCACCTTCCATCCAGACAGCCGAGCTCGCTCCGGAAC 690
 Db 804 GASGLGLSLGDFDGGGQ-----QGGDGSAGAGGAGGAGGAGGAGGAGGAGGAGG 846
 QY 591 TGTCTATCTC-----CAACGACAAAAG-----GACCTTTTGCCTGACTCGCGCTGT 738
 Db 847 GGDGATCAAGLNDGNGVGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 906
 QY 739 C-----CTGACACCTCCAGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 777
 Db 907 AGDNFNGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 966
 QY 778 CTTGGGAGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 825
 Db 967 GTGGDGGDAGAGG-----GGFGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1016
 QY 826 ----GGCCCAACTTCCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 881
 Db 1017 AGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1054
 QY 882 CTTGCTGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935
 Db 1055 TLGGPAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1111
 QY 936 GCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 995
 Db 1112 DPSIDG-----OGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1146
 QY 996 CTTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1041
 Db 1147 -----GNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1190
 QY 1042 ATGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1098
 Db 1191 -AGNAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1243
 QY 1099 CTTGATTTATGAACTCT-----ATAATAGATATATAGGTTT 1136
 Db 1244 GTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1303
 QY 1137 GTACCTTTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1196
 Db 1304 G-----GTNFGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1358
 QY 1197 AAAAG 1201
 Db 1359 GGLGG 1363

RESULT 7

Q8VIZO PRELIMINARY; PRT; 1715 AA.
 AC Q8VIZO;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE PE-PGRS family protein.
 GN MT3612.1.

QY 552 CTCAGCTTACCGAGTGGAACTCTACAGCGGTCTACGACTCATCTCTGACCTGCGAG 611
 Db 747 SGGAGTGGGAGGAGGAGADNPTGIGTG-GDGTGGAAGAGAG--GAAGTGTGTG 802
 QY 612 TAGTCTGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 671
 Db 803 MGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 845
 QY 672 CCGAGTCTCTCGGAACTTGCATCTCAAGCAAAAGAGAGGAGGAGGAGGAGGAG 731
 Db 846 AGAGGGGSSCAGTNGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 895
 QY 732 GCGGT----GTCTGACACCTCCAGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 787
 Db 896 GVGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 938
 QY 788 CCGGAACTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 834
 Db 939 FNGG-----TGCTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 981
 QY 835 TTGCGCTCTCCACTTGAATTCACCAAAATCCCTTCTGAGAGGAGGAGGAGGAG 892
 Db 982 -----AGDAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1012
 QY 893 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 945
 Db 1013 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1066
 QY 946 CAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1004
 Db 1067 --AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1109
 QY 1005 GCGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1063
 Db 1110 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1167
 QY 1064 GGTCTCTGCTCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1123
 Db 1168 GG-----AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1218
 QY 1124 GTATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1179
 Db 1219 G-GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1277
 QY 1180 TATATTAAGTATTAAGT 1202
 Db 1278 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG

RESULT 4

Q8VIY9 PRELIMINARY; PRT; 1217 AA.
 AC Q8VIY9;
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
 DE PE PGRS family protein.
 GN MT3615.3
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeGoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Deicher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBSJ databases.

DR EMBL; AB007164; AAK47974.1; --
 DR TIGR; MT3615; --
 DR InterPro; IPR002952; Eggshell.
 DR InterPro; IPR00084; PE region.
 DR InterPro; IPR002173; PfkB.
 DR Pfam; PF00934; PE; 1.
 DR PRINTS; PR01228; EGGSHLL.
 DR ProDom; PD01223; PE region; 1.
 DR PROSITE; PS00583; PFKB_KINASES_1; 3.
 SQ SEQUENCE 1217 AA; 94463 MW; 2F5BDCA5A30A056E CRC64;

Query Match 13.7%; Score 1035; DB 16; Length 1217;

Best Local Similarity 33.1%; Pred. No. 4.1e-53;
 Matches 409; Conservative 18; Mismatches 512; Indels 298; Gaps 48;

QY 1 GATCTGGGTCTCCAGGAAAGCAATCTCGA--AGTTAATGTTTG--ACTGAT 56
 Db 208 GGLTWGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 267
 QY 57 TTTTAAATCTCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 116
 Db 268 GGLWAPGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 303
 QY 117 TTGAATCCGCGGTCTTCGCGCTCAGACCCAGGAGGAGGAGGAGGAGGAGGAGGAG 176
 Db 304 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 352
 QY 177 TTAAGCGG--GCTGTGAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 231
 Db 353 CGDGTGSGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 412
 QY 232 TTGAATAAGAGGAGG--TGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 289
 Db 413 GTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 471
 QY 290 GCGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 349
 Db 472 ---GIGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 522
 QY 350 CTGAGTCACTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 407
 Db 523 GAGAGAGA-----DQFATGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 572
 QY 408 TGTCTCTCTGTGGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 467
 Db 573 QGGG-----AGSSGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 620
 QY 468 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 526
 Db 621 DGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 672
 QY 527 -----CTGGT-----ACCGGAGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 575
 Db 673 ADQFATGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 732
 QY 576 TACAGCGGTCTCATCGACTCATCTCGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 635
 Db 733 TGIGGTG-GDGGTGAAGA-----GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 775
 QY 636 GACCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 690
 Db 776 GA-----GGGG-----DGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 808
 QY 691 TGTCTATCCAAAGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 750
 Db 809 -----CGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 850
 QY 751 AGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 810
 Db 851 -----GNGGTGTGTGGV-----GTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 899
 QY 811 CG-----GACGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 866

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:03:25 ; Search time 57 Seconds
(without alignments)
6659.095 Million cell updates/sec

Title: X69111

Perfect score: 7544

Sequence: 1 GATCTGGGTGTCGCCAGG.....TTAACTTTTATAAAGTT 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 11 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1268.5	16.8	2147	16 Q98MG8	Q98MG8 rhizobium 1
2	1230	16.3	3145	16 Q98MG7	Q98MG7 rhizobium 1
3	1074.5	14.2	1489	16 O53559	O53559 mycobacteri
4	1035	13.7	1217	16 Q8VY9	Q8VY9 mycobacteri
5	1016	13.5	1938	16 Q7TWC0	Q7TWC0 mycobacteri
6	998.5	13.2	1460	16 Q7TWC3	Q7TWC3 mycobacteri
7	973	13.0	1715	16 Q8VIZ0	Q8VIZ0 mycobacteri
8	971.5	12.9	1079	16 O53557	O53557 mycobacteri
9	971.5	12.9	2639	5 O76786	O76786 antheraea p
10	970.5	12.9	2655	5 Q964F4	Q964F4 antheraea y
11	930	12.3	2395	5 Q27167	Q27167 paramecium
12	929.5	12.3	992	16 Q7TWB8	Q7TWB8 mycobacteri
13	894.5	11.9	2721	5 O76973	O76973 paramecium
14	869.5	11.5	1384	16 Q8VIZ1	Q8VIZ1 mycobacteri
15	867.5	11.5	2717	5 Q94710	Q94710 paramecium
16	864.5	11.5	1360	16 Q7TWC4	Q7TWC4 mycobacteri

17 863 11.4 988 5 O17434
18 860.5 11.4 1574 11 O88281
19 860 11.4 1381 16 O53552
20 852 11.3 1306 16 O53775
21 848 11.2 1306 16 Q7U1Q7
22 817 10.8 1507 16 Q8VJ23
23 807.5 10.7 1468 5 Q9GUB5
24 800.5 10.6 1665 16 O53215
25 797 10.6 1538 16 O53395
26 793.5 10.5 1408 16 O8VK17
27 789.5 10.5 773 16 Q7U160
28 789 10.5 1329 16 O6810
29 783.5 10.4 1246 4 O75095
30 779.5 10.3 1408 16 Q7U022
31 772 10.2 749 16 O53844
32 747 9.9 909 16 Q7U1D3
33 739.5 9.8 13288 6 O18758
34 726.5 9.6 1150 16 Q7TYG8
35 726 9.6 868 16 Q7U2D8
36 717 9.5 882 16 O53845
37 716 9.5 2189 5 Q9BI05
38 712 9.4 1664 5 Q9TVK2
39 712 9.4 2360 5 Q7Y2P0
40 709.5 9.4 879 16 Q8VKD2
41 706.5 9.4 853 16 O53439
42 701 9.3 850 16 Q7U0P1
43 687.5 9.1 837 16 O53684
44 685 9.1 831 16 Q7U2D6
45 684.5 9.1 2352 5 O61240

ALIGNMENTS

RESULT 1

Q98MG8 PRELIMINARY; PRT; 2147 AA.

ID Q98MG8
AC Q98MG8
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical glycine-rich protein mlr0585.
GN MLR0585.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idessawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002995; BAB48145.1;
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 27.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2147 AA; 181047 MW; FE2191AA8408E9AB CRC64;

Query Match 16.8%; Score 1268.5; DB 16; Length 2147;

Best Local Similarity 35.6%; Pred. No. 1.4e-66;
Matches 460; Conservative 64; Mismatches 628; Indels 141; Gaps 58;

QY 1 GATCTGG--GTGCTGCCAG-GAAACCAATCTG-----GAAGTAATGGTTTGA- 51

DB 702 GATGTGATGTGTGATGTGTGATGTGTGATGTGTGATGTGTGATGTGTGATGTGTGAT 761

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:51:29 ; Search time 18 seconds
(without alignments)
3480.022 Million cell updates/sec

Title: X691111

Perfect score: 7544
Sequence: 1 GATCTGGGCTGCCAGGA.....TTAAACTTTTATAAAGTT 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	970.5	12.9	1901	1 YZ08 MYCTU	O53553 mycobacteri
2	962.5	12.8	5263	1 FB0H_BOMBO	P05790 bombyx mori
3	893.5	11.8	2715	1 G156_PAPER	P13837 paramecium
4	857	11.4	2704	1 G168_PAPER	P17053 paramecium
5	784.5	10.4	914	1 WAZ2_MYCTU	O06794 mycobacteri
6	714.5	9.5	957	1 YZ78 MYCTU	P56877 mycobacteri
7	676.5	9.0	2531	1 NTC1_RAT	Q07008 rattus norv
8	674.5	8.9	2556	1 NTC1_HUMAN	P46531 homo sapien
9	670.5	8.9	5376	1 ZAN_MOUSE	O88799 mus musculu
10	668	8.9	778	1 YQ34 MYCTU	P71933 mycobacteri
11	666	8.8	481	1 LQRI_MOUSE	P18165 mus musculu
12	661.5	8.8	2531	1 NTC1_MOUSE	Q01705 mus musculu
13	656	8.7	2321	1 NTC3_HUMAN	Q9um47 homo sapien
14	654.5	8.7	1700	1 BAR3_CHITE	Q03376 chironomus
15	640.5	8.5	2319	1 NTC3_RAT	Q9r172 rattus norv
16	638.5	8.5	801	1 Y747 MYCTU	O53810 mycobacteri
17	637.5	8.5	2318	1 NTC3_MOUSE	O61982 mus musculu
18	631	8.4	747	1 SPDI_NEACL	P19837 nephila cla
19	616.5	8.2	2437	1 NTC1_BRARE	P46530 brachydanio
20	603.5	8.1	2524	1 NOTC_XENLA	P21783 xenopus lae
21	603.5	8.0	672	1 PHX5_MOUSE	F08399 mus musculu
22	588	7.8	2703	1 NOTC_DROME	P07207 drosophila
23	583	7.7	2471	1 NTC2_HUMAN	Q04721 homo sapien
24	579	7.7	1964	1 NTC4_MOUSE	P31696 mus musculu
25	577.5	7.7	1046	1 PSTA_DICDI	P11376 dictyosteli
26	577.5	7.7	2471	1 NTC2_RAT	Q9gw30 rattus norv
27	574	7.6	2470	1 NTC2_MOUSE	O35516 mus musculu
28	557	7.4	4289	1 TENX_HUMAN	P22105 homo sapien
29	540	7.2	2907	1 FBN2_MOUSE	O61555 mus musculu
30	535.5	7.1	2003	1 NTC4_HUMAN	Q99466 homo sapien
31	535	7.1	1064	1 FBN1_STRPU	P10079 strongyloce
32	527	7.0	2911	1 FBN2_HUMAN	P35556 homo sapien
33	514.5	6.8	641	1 EBN1_EBV	P03211 epstein-bar

ALIGNMENTS

RESULT 1

YZ08 MYCTU
ID YZ08 MYCTU STANDARD; PRT; 1901 AA.
AC Q53553;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv3508 precursor.
GN Rv3508 OR MV023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37Rv;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekoa P., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R.T., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544 (1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AL022022; CAA17745.1;
CC PIR; F70806; F70806.
CC TuberculList; Rv3508.
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.
CC ProDom; PD001223; PE_region; 1.
CC Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN Rv3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match 12.9%; Score 970.5; DB 1; Length 1901;

Best Local Similarity 31.9%; Strd. No. 1.3e-42;

Matches 405; Conservative 19; Mismatches 527; Indels 319; Gaps 48;

OY 1 GATCTGGGCTGCCAGGAAGCAAAATCTGGA--AGTTAATGTTTTC--AGTGAT 56

[illegible]

GN 156G.
 OS Paramesium primaurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Paramesium.
 OX NCBI_TaxID=5886;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=156;
 RX MEDLINE=87060934; PubMed=3783679;
 RA Prat A., Katinka M., Caron F., Meyer E.;
 RT "Nucleotide sequence of the Paramesium primaurelia G surface protein.
 ET A huge protein with a highly periodic structure.";
 EL J. Mol. Biol. 189:47-60(1996).
 CC -!- FUNCTION: This protein is the surface antigen or immobilization
 CC antigen of Paramesium primaurelia.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DOMAIN: It has internal homologies and a highly periodic structure
 CC with 37 periods of about 75 residues, each period containing 8
 CC cysteines, except for four half periods. A variable part of 475
 CC residues comprises 4 almost identical periods in the middle of the
 CC protein.
 CC -!- MISCELLANEOUS: Expression of G protein occurs at low temperatures
 CC (14-32 degrees Celsius).
 CC -!- SIMILARITY: Contains 34 PSA repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X03882; CAA27514.1; -;
 CC PIR; A23475; A23475.
 CC InterPro; IPR002895; Paramesium SA.
 CC Pfam; PF01508; Paramesium_SA; 34.
 CC SMART; SM00639; PSA; 33.
 CC Signal; Repeat; Antigen; Membrane; GPI-anchor.
 KW SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2715 G SURFACE PROTEIN, ALLELIC FORM 156.
 FT DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.
 SQ SEQUENCE 2715 AA; 279551 MW; 97BE359AB9C7C298 CRC64;
 Query Match 11.8%; Score 893.5; DB 1; Length 2715;
 Best Local Similarity 24.6%; Pred. No. 1.3e-38;
 Matches 453; Conservative 56; Mismatches 630; Indels 701; Gaps 102;
 QY 3 TCTGGGGTCTGCCAGGAAAAGCAAAATCTGGAAGTTAATGTT-----TTTGAG 52
 DB 835 TCTVYKVGAGCC-----VKSANGSDYMTSAQCHKTLTNLTANDCKWIVDRCYALSSFA 890
 QY 53 TGATTTTAAATCC---TTGCTGGCGGAGAGCCGCTCTCCCGGTATCAGCGCT--T 107
 DB 891 TGACTTFKGTMCYRAGCTNTVGAASSASCTLDCTLT-----GSLTFAD 939
 QY 108 CTTATTTTGAATCCCGGCTCCGGGCTTCGGGTGAGACAGCCGAGGAGCGCT 167
 DB 940 CQALDSTGCVKDGTT-----GCATQSTCAGYGSTAAACFR---SSASGTAGYC- 985
 QY 168 GTTTCGAATTAAGCG--GGGTGTG-----AACGCCAGGCC-----CGCGGGGG 211
 DB 986 AMNTNCQSVTSAEACAFVTLGLDHSKQLYHSSCTSLKDGTCGCEYKTTCSGYAATNN 1045
 QY 212 C---GGGCGGAGGC---GGGCCATTTGAATAAGA---GGC-----GTGC---CTTCCAG 256
 DB 1046 CATSGGKCFDFVECLRFNSCASITGTGLTTA;CGTYDAGCVANVNGTACQEKLATCDLY 1105
 QY 257 GCAGGCTCTATAAGTGACC-----GCCCGGCGAGCGTGCAGCGGTTGCGAGGTCACT---- 308
 DB 1106 LTQNSCSFSAAATADKCAWSGTAFLAVTTVGTGHPYVGTGLTDLICAYNANCANKA 1165
 QY 309 GTAGCGGACTCTTTTGGTTTCTTCTTTTGGGCACTCTGGAC-----TCATCTCCC 364

DB 1166 GTA-CQEKATCNLYT--TEATCSTSAATAADK-----CAMSGAACLAVTTVATECAV 1217
 QY 365 AGCATG-----AAGCGGTG--AGCCCGTGGGGGTGCTAGCAGGGCGGTGCTGCC 416
 DB 1218 TG--TGLTDLICAYNANCANKAGTACQEKATCNLYTTEATCSTSAATAADK----- 1271
 QY 417 TGTGGAACGCACTCTGGCCATCGCCCGGGCGGAGGAGGCGCGGCGAGCTGAGGAGC 476
 DB 1272 --AWSGAACLAVTTVATECAVYTGTLTNAIC-----AAYNANCANKAGTACQEKAT 1323
 QY 477 CG--CTGAGTCTGCTGAGACA-TGAACCA-----C-----TGCTACTGCC--GCCT 519
 DB 1324 CNLYTTEATCSTSAATAADKCAWSGAACLAVTTVATECAVYTGTLTNAICAYNANC 1383
 QY 520 GCGGAACTGGTACC-----CGGAGTCCGAGGAGGCACTCAGC-----TTAGCCA--G 565
 DB 1384 ANKAGTACQEKATCNLYTTEATC-----STSAATAADKCAWSGAACLAVTTVATECAV 1440
 QY 566 GTGGAATCCTACAGC-----GGTCACTGACTACATT----- 598
 DB 1441 GTGLTTRAIATYVAGCINLKDTGCGEAKANKDYTTTSNKTAQTITSLCLWIDNSCYP 1500
 QY 599 -----CT-----CGACTCGAGTATGCTCTG--GCGG--AGCCAGC-----C 631
 DB 1501 VTDLNCSTVITGLGFVHAQCOAVSTGCTSVSDGSKQDFKSTCEQYPTGLGCTKTATKC 1560
 QY 632 CCTGGAC-----CCCTGTATG-----GC-----CC-----C--CA 654
 DB 1561 YLQGSACITISNATDCAKITSGAGTITPEICQSVNTGCVNRSACVQQAQCSGYTS 1620
 QY 655 CCTTC-----CCA-----TCAGACAGCC-----GAG-----CT---CGCTCCGAA 688
 DB 1621 AMTSCYKSGAGLCIASNTDITACVAATAATDAVILGAGNYSSANCNEMKAGCTNNGTA 1680
 QY 689 CTGTGTATC-----TCAACGACAAAGAGAGCTTTTGCCACTGACTC-----GGCG 735
 DB 1681 CVAKTCANAAGITFNHTNCNLYNTGNSGSAQCTWASKADQTOASCLYSVEGEV 1740
 QY 736 TGTCTT-----GACT-----ACCTCCAGAAC--GCA 758
 DB 1741 VGTSCVRKTCDAATDATTDDTECTYQCSCTVARLGACQARAACATYKSLQCKFNTS 1800
 QY 759 GG-----TGCTGGCGCCCGTT-----CTGCTGGGAGCCCGGAA 793
 DB 1801 GSKCFWNPINKTVDLNCNGIEATTLTDHNECAVDATLACTVETATNGAAQGCWARGA 1860
 QY 794 CCTCT-----CCTCCCG--AAGCCGAC-----GGC----- 818
 DB 1861 CASYTTIEBQCKTNASGVCMVNTNANLPAPACQDKSCTSAPTSTTTTHNDYAYNTATVK 1920
 QY 819 -----AGGATGGGCCCAACTT-----CG-----CCTGCG----- 844
 DB 1921 CTVVATPSNSGNPTLGGCCQTAACSSYIDKECCQINANGDPGWNGTCCADKSCATASA 1980
 QY 845 -----CCACTGACT-----TC-----ACCAATCCTC 867
 DB 1981 TADYDDTKRAYITNKCTVSDSGGCVETPATCTMTQKOCYNKAGDPCTWTGTACIT 2040
 QY 868 TCTGGAGA-CTAAACCTGTGCT-----C-----AGGAGCGAAGGACTGTGAAC 911
 DB 2041 KSCDNAPDATATADENCTYLAGCTLNNVKCKTKVCDFAFATDALCKQAICTCTINGNC 2100
 QY 912 TT-GT-----AGCCTGAGAGC-----CAGAGC-TAGCTCTGCC 944
 DB 2101 VTGTCFQALSOAGCVTSSTNOCCWIPAVLNASNVITSPAYCTTKNGCTAPITLISE-- 2158
 QY 945 ACCAG-----CT---GGCGAGCTCACCCCTGCTCCCAACCCCAAC-----CCCAA 985
 DB 2159 AACAGYFTNCTTKNGGCVTKSTCSAVTIDVACTTALNGTVCWDSAQNKCRDKDCQDFS 2218
 QY 986 GTT-----CTA-AGGTCTTT-----TCAGAGC--GTGG-----AGGTGTGA 1019

Db 2219 GTTHAACQAGACGTAGAGGKCARVQNCQTSVRAACIEGTNGPCLWIDKYQNTDGTGA 2278
Qy 1020 A-----GGAGT-----GGC----- 1028
Db 2279 CFRTYCKSLNWNDSCKWISNCKTNGNCVIGITLCSBTNDGCVTVYDCACTQSV 2338
Qy 1029 -----TGCT-----CTCCAAACT-----ATGCCAAGCGGC-----GGC----- 1057
Db 2339 DLNSSDPKVKPYTSCADAFYTHSDQIASSCKTNGTGTGIALGSCSYTVQAGCVFN 2398
Qy 1058 -AGAGCTGGTCTTCTGCTCTCCITGGAGAAAGTCTGTGGCCTGATTTAGAACTC-- 1114
Db 2399 DKGTLTSGVITST--GICTWDTTSSCRDQSCADLTGTTHATCSSQLSTCTSDGTTCLL 2456
Qy 1115 -----TATAATAGAGT-----ATAT-----AGGTTT----- 1135
Db 2457 KGACTSYTTQACTTAVGSDGACWELASATNNNTAKRLLTCAQLQNGTATNVCSVALS 2516
Qy 1136 -----TGATAC-----CTT-TTTTAC-----AGGAGTGACT--TTCT--- 1165
Db 2517 TCVSNGTACIPKANCSTYTSKVACNSGGLDGI CVFTQSTATGAAGTGTCTALMTACTVAN 2576
Qy 1166 -GTAACAA-----TGCGATGTATATAAACTTTTATAAA 1199
Db 2577 NDQTAQCAARDRCSTWTAASGTRATAVASKATHCATNOA 2616

RESULT 4
G168 PARPR
ID 168 PARPR STANDARD; PRT; 2704 AA.
AC P17053;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE G surface protein, allelic form 168 precursor.
GN 1688.
OS Paramacium primaurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramacium.
OX NCBI_TaxID=5886;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=90172419; PubMed=2308165;
RA Prati A.;
RT "Conserved sequences flank variable tandem repeats in two alleles of
the G surface protein of Paramacium primaurelia.";
RL J. Mol. Biol. 211:521-535(1990).
CC -!- FUNCTION: This protein is the surface antigen or immobilization
antigen of Paramacium primaurelia.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- DOMAIN: It has internal homologues and a highly periodic structure
with 37 periods of about 75 residues, each period containing 8
cysteines, except for four half periods. A variable part of 475
residues comprises 4 almost identical periods in the middle of the
protein.
CC -!- MISCELLANEOUS: Expression of G protein occurs at low temperatures
(14-32 degrees Celsius).
CC -!- SIMILARITY: Contains 34 PSA repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52133; CAA36378.1; -;
DR PIR; S09118; S09118.
DR InterPro; IPR002895; Paramacium_SA.
DR Pfam; PF01508; Paramacium_SA; 34.
DR SMART; SM00639; PSA; 33.

KW Signal; Repeat; Antigen; Membrane; GPI-anchor.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2704 G SURFACE PROTEIN, ALLELIC FORM 168.
FT DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.
FT DOMAIN 1060 1424 VARIABLE DOMAIN, COMPRIS 4 ALMOST
IDENTICAL REPEATS.
SQ SEQUENCE 2704 AA; 278775 MW; 40EAOA0B18EE2119 CRC64;
Query Match 11.4%; Score 857; DB 1; Length 2704;
Best Local Similarity 23.8%; Pred. No. 9.4e-37;
Matches 440; Conservative 59; Mismatches 630; Indels 716; Gaps 98;
Qy 3 TCTGGGCTGTCGACGAAAGAAATCTGGAAGTTAATGCT-----TTTGAG 52
Db 829 TCTVYKVGAGC-----VKSANCSYMTSAQCHKLTNLNTANDCKWIVDRCYALSSFA 884
Qy 53 TGATTTTAAATCC-----TTGCTGGCGAGAGCCCGCTCTCCCGGTATACGCT--T 107
Db 885 TGACTTFFKGNKWEYGRAGCTNTVGAASASCTLDCTLKT-----GSLTTFAD 933
Qy 108 CCTCATTTTGAATCGCGGCTCCCGGCTTTCGGCGTTCAGACCAGCGGAGGAGCT 167
Db 934 CQALDSTCSVKDGTGCVIQTSCAGYGSTATNC-----FRSSASGTAGYC- 979
Qy 168 GTTGGCAATTTAAGC--GGCTGTG-----AAGCCCGAGGC-----CGGCGGGG 211
Db 980 AMNTNQSVTSAACAFVTGLTGLDHSKCOLVHSSCTSLKGTGQEQYKTCSSYATGNT 1039
Qy 212 CGGG--GCC--GAGGC--GGGCCATTTTGAATAAAGA--GGC-----GTGC----- 249
Db 1040 CANSVQKCPDDATDCLRFANCASTTGTGLTNTICVTYDPGCVANVNGTACQEKLATCAA 1099
Qy 250 -----CTTCCAGCA--GGCTC-----TATAAGTACCGCGCGCGGAGC-----G 288
Db 1100 YLTQNSCTSTAGTCAWSGACLTVDVANVATECAVITGTGLTNAICAGYNAKCTVNRAG 1159
Qy 289 TGC-----CGCGTTGCGAGTCACTGTAGC--GGACTTCT-----TTTGTGTTTC 331
Db 1160 TACQKEALCATYAAVQATCSQSDAGLCAWSGACLTVDVANVATECPYITGTGLTNAIC 1219
Qy 332 T-TTCTTTTGGCGGACCTCTGACCTACTCCCGCATGAGGCGCTGAGC----- 382
Db 1220 AGYNAKCTVNRAG-----TACQKEALCATYAAVQATCSQSDAGLCAWSGACLTVDVAN 1275
Qy 383 ---CGGTGCG-----CGG-----CT-----GCTACGAGGCGG 407
Db 1276 ATGCTPYITGTGLTDAICAGYNAKCTVNRAGTACQKEALCATYAAVQATCSQSDAGLCAW 1335
Qy 408 TGTGC-----TGCTGTGCGAAGCAGCTCTGGCC---ATCG----- 440
Db 1336 SGACLTVDVANVATECPYITGTGLTNAICAGYNAKCTVNRAGTACQKEALCATYAAVQ 1395
Qy 441 --CCCGGGCGGAGGAGGAGGCG-----CCGCGAGCTGAG--GAGCGG---CTGAGCT 485
Db 1396 ATCSQSDAGLCAWSGSA---CLTVVDVANVATECAVITGTGLTDAICAGYNAKCTNLKDG 1451
Qy 486 TGC-----TGAGCAGCATGAGC-----ACTGCTACTC-----CCGC 517
Db 1452 TGQDEKATCKLVTYQNKCTSTQTTPLSLWFDNSCSPTDVTCSAIVOSGLDHAQCAQY 1511
Qy 518 CTGC---GGGAAGTGGTACCGGAGTCCCGAGAGGAGCAGCTAGCTTAGCAGGTGGAATC 574
Db 1512 STGCTSVSDGSKQDFKTCQVAGT-----ALSTKTATSKCYLQ-----SNC 1556
Qy 575 CTACAGGCGTCTACCTACATCTCT---CGACCTGAGGTAGTCT---GGCGGAGCCAG 629
Db 1557 ITI-----SNVATDCAKITGSGITTYEIQSINTGCSVNRARSACVQQAQCSGVTSM 1611
Qy 630 CCCTGTGAGCCCTGTAGTGGCCGCCACCTTCCCATCCAGACCGAGCT---CGCTCCGG 686
Db 1612 TSCVKGAGLCLASTNTDTACVAATAATTCDAVYLTG---NYSSANCNEMKAGCTNNGA 1668
Qy 687 AACTGTCTATC-----TCCACAGACAAAGGAGCTTTTGGCCACTGACTC-----GGCC 734

Db 1669 TACVAKTCANAVIFNHTNCGYLNCTCTVNSGACQTMASKADQTOASCLYSVEGCV 1728
 Qy 735 GTGTCC-----TGA-----CAC-----CT-----CCAGAAGC-----CAGGT 761
 Db 1729 VVGTCVKTCTAATDTRDDTSCSAYQSCCTVARLGCQARACASYSKSLQCKENT 1788
 Qy 762 GCTGG-----CCCCCGTT-----CTGCCTGGACCCCGG 790
 Db 1789 --SGRCFNPNKTCVDLNCNIEASTLYDTHNECVVVDATLACTVATNGAAVQCGMA 1846
 Qy 791 GAACTCT-----CTGTCCCG-----AAGCGGAC-----GGC-----818
 Db 1847 RGACSSYTBEOCKTNASGVCMVNTNANLPAQODKSCTSAPTSTTHNDCAVYNTA 1906
 Qy 819 -----AGGATGGGCCCCAAGTT-----CG-----CCCTGC-- 844
 Db 1907 TVKCTVAFPSNGNPTLGGCQQAACSSYIDKQCOINANGPCGWNQTQADKSCAT 1966
 Qy 845 -----CCACTTGACT-----TC-----ACCAATC 864
 Db 1967 APATADYDDTKRAYITNKCTVSDSGQGVPEIPATCTMTKQCYVKNAGDPCTWTGTA 2026
 Qy 865 CTTCTCTGGAGA-CTAAACTGTGCT-----C-----AGGAGCGAGGACTGTG 908
 Db 2027 CITKSCDNAPDATADECNCTYLAGCTLDNVKTKVCEDFAPATDALCKQAISTCTTNG 2086
 Qy 909 AACTT-GT-----AGCCTGAAGAC-----CAGAGC-TAGCTGTG 941
 Db 2087 TNCVTRGTCPQALSOAGCVTSSTNQCEWIPAVLNASVITSPAYCTIKNCSTAPITLTS 2146
 Qy 942 GCCACGAG--CT-----GGCGGAGCTCACCTGTCTCCACCCACAC--CC 982
 Db 2147 E-AACAGYTNCTTNGGCVTKSTCSAVTIDVACTTALNGTVCAWDSQNKCRDXCQ 2204
 Qy 983 CAAGTT-----CTA-AGTCTTT-----TCAGAC--GTGG-----AGGTGT 1016
 Db 2205 DFGTTHAACQACQACGTAGAGKCARVQNCQTSVRAACIEGTNGPCLWIDKYQNTDGT 2264
 Qy 1017 GGAA-----GGAGT-----GGC-- 1028
 Db 2265 KGACFRYTSKSLNWNDSCKWISNCTYNGCNVGLCSFTNTDGGCVTCYDGAQIQ 2324
 Qy 1029 -----TGCT-----CTCAAACT-----ATGCCAAGCGCGGCAGA 1060
 Db 2325 SVPALNSSDPKVPYTSKADAFYTHSDQIASSKCTTNGTTCIALGSCSSY--TAQA 2382
 Qy 1061 GC-----TGCTCTCTGCTCTCTGGAGAAGTTCTCTGCTCTGATTTATGAA 1111
 Db 2383 GCYFNDKGLTYSVTST--GICWTDTSSCRDQSCADLTGTTATCSQLSTCTSDG 2440
 Qy 1112 CTC-----TATAATAGAT-----ATAT-----AGGTTT-- 1135
 Db 2441 TTCLLKGACTSYTTQTACTTAVGSDGACVWELASATNNNTAKRLLTTCADIQNGTATNVC 2500
 Qy 1136 -----TTAC-----CTTTT-----TTACGGAAGTGACT--TT 1163
 Db 2501 SVALSTCVNSGTACIPKANCSTYTSKACNSGGLDGCIVTQTSTATGAAGTGTALMTA 2560
 Qy 1164 CT-----GTAACAA-----TGCGATGTATATAACTTTTATATAA 1199
 Db 2561 CTVANNDQTAQOARDRCSTWTAASGTGATAVASKATHTCATNQA 2605
 RESULT 5
 ID WA22 MYCTU STANDARD; PRT; 914 AA.
 AC O06794;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE WAG22 antigen precursor.
 GN WAG22 OR WAG22B OR RV1759C OR MT1907 OR MTY28.25C OR MB1789C.

OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."; Nature 393:537-544 (1998).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=2206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., S.L.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains"; J. Bacteriol. 184:5479-5490 (2002).
 RL [3]
 RN SEQUENCE FROM N.A.
 RP SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Aikin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RA "The complete genome sequence of Mycobacterium bovis";
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 RL [4]
 CC -|- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
 CC SUBFAMILY.
 CC -|- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to a
 CC frameshift in position 85.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; 295890; CAB09322.1;
 DR EMBL; AE007040; ; NOT ANNOTATED CDS.
 DR EMBL; BX248340; CAD94491.1; ALT_FRAME.
 DR PIR; H70987; H70987.
 DR TIGR; MT1807; ;
 DR TubercuList; RV1759C; ;
 DR InterPro; IPR00084; PE_region.
 DR Pfam; PF00934; PE; 1.
 DR Antigen; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 914 WAG22 ANTIGEN.
 SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBE8B6AC8 CRC64;
 Query Match 10.4%; Score 784.5; DB 1; Length 914;
 Best Local Similarity 29.5%; Pred. No. 1.9e-33;
 Matches 337; Conservative 18; Mismatches 386; Indels 401; Gaps 48;
 Qy 1 GATCTGGGTGCTGTCGAC-----GAAAAAGCAAAATCTCGGAAGTTAATGTTTTCAGTGA 55

Db	116	NGANGAFTGANGGDAGWLLIGNGAGSGAKGA---NGAGGPGGAAGLFENGAGG-GA	171
QY	56	TTTTTAAATCCCTGCTGGCGGAGAGGCCCGCCCTCTCCCGGTATCAGCGCTTCCTCATTC	115
Db	172	GGTATA-----NNGIGGAGG-----AGGSAMLFGAG-----	197
QY	116	TTTGAATCCGGGCTCGCGGTCTTCGCGTCAAGCAG-----CGGAGGAGACCT	167
Db	198	---GAGGAGGAATSLVGGIGGTGTGNAGMLAAGAGAGAGGFFSFTAGAGAGG---	251
QY	168	GTTTCAATTTAAGCGGCTGTGAACCCAGG--GCCGCGGGGCGGGCCGAGCGG	225
Db	252	---AGGLFTTGGVGGAGGCGGHTGAGGAGGAGLFG-AGMGAGAGGFGDHGTGLTGTGAGG	307
QY	226	GCCATTTTGAATAAAGAGGCGTGCCTTCAGGCGAGGCTATATAAGTGAACCGCGCGGCGA	285
Db	308	-----DGGGGCLFAGG-----DGGAGGSLTTGGAGNGGAGTILSLAAG-	349
QY	286	GGGTGGCGCGGTGCAGGTCACTTAGCGGAAGTCTTTTGGTTTCTTTCTTTTGGGCG	345
Db	350	-----GAGGTGAGGTFFGGKGGAGGA-----GGN-	375
QY	346	ACCTCTGGACTACTCCCCAGCATGAAGCGCTGAGCCGCTGCGCGGTGCT--ACGAG	403
Db	376	-----AGMLFSSGGGTGCGFPFAGGCG--GVGSAGMLSGSGSG	415
QY	404	CGCGTGTGCTGCTGTGGGA--ACGCAAGTCTGCCCATCGCCCGGGCGCGAGGGAA---	457
Db	416	GAGGSG-CPASTAAGGAGGAGGAPFLIG--NGNGNGGSESGTGGVGGAGGNAVLING	472
QY	458	GCCCCGGAGCTGAGGAG--CCGCTGAGCTTGC-----TGGACGACAT	498
Db	473	GSGGIGALAGKSGFGGFGGLLAGDYNAPESISPWHLQQDILSPINEPTALTRPLI	532
QY	499	GAACACATGCTACTCCCGCTCGCGGAAGCTGTGTACCGGAGTCCCGAGGAGCATCTAGCT	558
Db	533	GNDSGTPG-----TGDDEGA--GGWLFNGNGG---GAGAAGTNGSAGGAG	574
QY	559	TAGCAGAGTGGAAATCCTACAGCCGCTCATCGACTACATTCTCGAGCTCGAGGTAGTCTCT	618
Db	575	GAGGILFTGGAG-----GAGVGGTAGA-----GGAGGAGGSNAFL	609
QY	619	GGCCGAGCCAGCCCTGGACCCCTGTATGCCCCCACCCTTCCATCCAGACAGCGAGCT	678
Db	610	ISGGTGGVGGAAFTTGG-----VGAGG-----NAGLLIGNAAGL--	645
QY	679	CGCTCCGGAATCTGTTCATCTCCAACGACAAAGAGGCT-TTTGCCACTGACTCGCGCCGTG	737
Db	646	---GCGGGAFTAG-----VTTGAGGTGGAAGLPANGA--GGAGTG	683
QY	738	TCTTGACACTCCAGAAGC--CAGGTGCTGGCGCCGTCTGCTCGTGGAGACCGGGAACC	795
Db	684	STAGGAGG---AGGAGGLYAHGGTGGPGNGGSGTGGGTGAGGPGAGGPGGLYAGGS---	734
QY	796	TCTCTCGCGGAAGCGGACCGCAGGAGTGGGCCCAACTTCGCGCTCCCACTTGACTT	855
Db	735	-----GGA-----GGHGGNAGGGGVGG-----NAGSLTL	759
QY	856	CACCAATCCCTTCTTGGAGACTPAACCTGGTCTCAGGAGGAGAGGACTGTGAAGTGT	915
Db	760	NA-----SGAG-----GSGGSSLSGKAGAGGAG-----GS	785
QY	916	AGCCTGAAGAGCCAGAGCTAGCTCTGGCCACCAGCTGGCGGACGCTCACCTGCTCCCAACC	975
Db	786	AGLFGYGGGAGNGYSLING---TGG-----DGGTGGAGQ-----	817
QY	976	CCACCCCAAGTCTAAGGTTCTTACAGCGTGGAGGTGTGGAAGGAGTGCCTGCTC	1035
Db	818	-----ITGLRGGP-GAGGAGGASDTGAGNGGAGGKAG-----	850
QY	1036	CAAACTATGCCAAGCGCGCAGAGCTGGTCTTCTGTCTCTCTTGGAGAAAGGTTCTGT	1095

Db 851 -----LYNGCGDGGAGGDNATSGKGGAGGNAVVING-----GNGGNACKAGGTTAGAG 899

Qy 1096 TG 1097

Db 900 AG 901

RESULT 6

ID	Y278 MYCTU	STANDARD;	PRT; 957 AA.
AC	F56877;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Hypothetical PE-PGRS family protein RV0278c/MT0291 precursor.		
GN	RV0278C OR MT0291 OR MT0355.06C.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]	SEQUENCE FROM N.A.	
RC	STRAIN=H37Rv;		
RC	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.K., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Sadcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtzoyd S.,		
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rejandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,		
RA	Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence."		
RL	Nature 393:537-544(1998).		
RN	[2]	SEQUENCE FROM N.A.	
RP	STRAIN=CDC 1551 / Oshkosh;		
RC	MEDLINE=22206494; PubMed=12218036;		
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,		
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,		
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,		
RA	Biswal W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;		
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and		
RT	laboratory strains";		
J.	Bacteriol. 184:5479-5490(2002).		
CC	-!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS		
CC	SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to licenses@isb-sib.ch).		
CC	-----		
EMBL:	AL021930; CRAI17353.1; --		
DR	PIB; AE006936; AAK44511.1; ALT_INIT.		
DR	PIB; D70835; D70835.		
DR	TIGR; MT0291; --		
DR	TubercuList; RV0278c; --		
DR	InterPro; IPR000084; PE_region.		
DR	Pfam; PF00934; PE; 1.		
KW	Hypothetical protein; Repeat; Signal; Complete proteome.		
FT	SIGNAL 1 30 POTENTIAL		
CH	CHAIN 31 957		
FT	HYPOTHEICAL PE-PGRS FAMILY PROTEIN		
FT	RV0278C/MT0291.		
FT	M -> I (IN REF. 2).		
FT	CONFLICT 158 163 MISSING (IN REF. 2).		
FT	CONFLICT 807 8105 R -> G (IN REF. 2).		
FT	SEQUENCE 957 AA; 81905 MW; 715BAD417FFA47C CRC64;		
SQ			

Query Match		9.5%; Score 714.5; DB 1; Length 957;
Best Local Similarity		30.1%; Pred. No. 6.8e-30;
Matches 352;		Conservative 19; Mismatches 406; Indels 393; Gaps 58;
QY	21	AAACCAAAATCTGGAAGTTAATGTTTGTAGTGAATTTTAAATCTTCTGCGGAGAG 80
DB	12	AAAATDLASLSSSIISANAAAAAANTTALMAAGADEVSTAAAL-----FGAH 58
QY	81	GCCGCCCTCTCCCGG---TATACGCGTCTCTCTCATTT-----CTTTG----- 119
DB	59	GOAYQALSAQAQAFQAFQVQLTSGGAYAAEAANVPLDPINFFLANTCRPLING 118
QY	120	---AATCCCGGTCGCGGTCCTTCCGCGTCAGACACCGGAGGAAAGCCCTGTTTCAATT 177
DB	119	ANGAPGTGANG-----GDGWLINGGAG--GSGAGVNGGAGNGG----- 158
QY	178	TAAGCG--GGCTGTGAACGCCACGCGCGCGCGCGCGG---GCCGAGCGCGGCATT 231
DB	159	---AGNGGAGGLINGGAGG---AGVASSGIGSGGAGGNAMLFAGAGGAGGAVVA-- 212
QY	232	TTGAATAAGAGCGGTGCTTCCAGGCGAGGCT---CTATAAGTACCGCGCGCGGAGCG 288
DB	213	LTGAGAGGAGGAG--NAGLLFGAAGVGGAGGFTNGSALGGAGGAGGGLFATGGVGGSG 271
QY	289	TGCGCGCTTGCGAGTCACTAGCGGACTTCTTTTGTGTTTCTTCTCTTTGGGCGACC 348
DB	272	ASSSGAGGAGAGGLFAGGTGHHGFADSSFVGVG----- 309
QY	349	TCTGACTCACTCCACCATGAAGCGCTGAGCCCGTCCGCGGCTGCTAG--AGCGG 407
DB	310	---AGG-----AGLFGAGGEGSGSHSVAGG--GAGNAGMALGAGAGG 355
QY	408	TGTGCTGCT--GTGGAACGCGAGTCTGCGCATCCCGCGCGCGGAGG-----GAAGGCGC 461
DB	356	IG-GDGLTLAGGIGA-----GGAGNAGLLFGSGSGGAGGFGFADGGQGP-- 403
QY	462	CGGAGCT--GAGGAGCGCTGAGTTG--TGAGAGCATGAACCACTGCTACTCCCGC 518
DB	404	---GNNAGTVFGSGGAGGNGVGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 445
QY	519	TGCGGAACTGGTACCCGAGTCCGAGAGCACTCAGCTTAGCCAG--GTGGAATCC 575
DB	446	GGNGASAVTGNG--GIGTGVLVINGNGSGSGIGAGKAGVGVSGLLGLDGNFAPAS 504
QY	576	TA-----CAG---CGGTATCATCACTACATTCGACCTGACGTA 613
DB	505	TSPLHTLOONLVNVNPFQTLGRPLINGANGTPG-----TGADGGAG 549
QY	614	GTCTGCGCGAGCCAGCCCTGAGACCCCTGATGCGCCGCCACCTTCCATCCAGACAGCC 673
DB	550	GWLFGANG-----TPG-----TGAAGG-----AGGWLFGNG 577
QY	674	GAGCTCGTCCGGAACCTGTATCTCCACGAAAGAGAGCTTTGCCACTGCTCGC 733
DB	578	GNG-----GHGA--TNTAATAT-----GGAGGAG-----GILFGT-----GNG 608
QY	734	CGTGTCTTGACACCTCCAGAACGAGGTGCTGGCGCCGCTTCTGCTGGACCCCGGAA 793
DB	609	GCTGGIATGA-----GGIGAGGAGVGLLIGSGGTG--GNGGNSIGVAGIGA 655
QY	794	CTCTCTCTCGGAGCGGAGC---GCAG-----GGATGCGCCCACTTCGCGCTG 843
DB	656	-----GGGGDAGLLFGAAGTGGGAAGVPAVGAGGAGG-----GLFANG 697
QY	844	CCACTTGACTTCACCAATCCCTTCTGAGACTAAACCTGTGCTCAGAGCGAAGA 903
DB	698	-----GA-----GGAGGFNA-----GGNGN--GGLFTGTGTGA 726
QY	904	CTGTGACTTGAGCTGAAGCCAGCTAGCTCTGCGCACCACTGGCGGCGACCTCAC 963
DB	727	GTNFGAG-----GNGGNGGLFGAGGTGGAAGSGG-----SGITTTGGGHHGNAGL 771
QY	964	CTGTCTCCACCCACCCCAAGTCTTAAGGTCCTTTTTCAGAGCGTGGAGGTGGA----- 1019
DB	772	LSLG-----ASG-----GAGGS-----CGASSLAGAGGTGGNALLFG 805
QY	1020	---AGGAGTGGCTGCTCTCAAACTATGCCAAGGCGGCG---GCAGAGCTGCTTTC-- 1070
DB	806	FRGAGGAGGAG-----GAALTSIQGGAGGAGGGLLFGSAGAGGAGGAGGAGGAG 856
QY	1071	---TGCTCTCTCGAGAAAGGTTCTCTTGC 1098
DB	857	ATGG-----TGDDGGHAGVFGNGGDDGC 880
RESULT 7		
ID	NTC1 RAT	
AC	Q07008;	STANDARD; PRT; 2531 AA.
DT	01-NOV-1995 (Rel. 32, Created)	
DT	15-JUL-1999 (Rel. 38, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Neurogenic locus notch homolog protein 1 precursor (Notch 1).	
GN	NOTCH1.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Euteleostomi; Chordata; Vertebrata; Euteleostomi;	
OX	NCBI_TaxID=10116;	
RN	[1]	SEQUENCE FROM N.A.
RP	TISSUE=Schwann cell;	
RX	MEDLINE=92111383; PubMed=1764995;	
RT	Weinmaster G., Roberts V.J., Lemke G.;	
RT	"A homolog of Drosophila Notch expressed during mammalian	
RT	development."	
RL	Development 113:199-205 (1991).	
RN	[2]	REVISIONS TO 1652-1653.
RP	Weinmaster G.;	
RX	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.	
RN	[3]	FUNCTION.
RX	MEDLINE=21094508; PubMed=11182080;	
RA	Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,	
RA	Honjo T.;	
RT	"Notch1 and Notch3 instructively restrict BFGF-responsive multipotent	
RT	neural progenitor cells to an astroglial fate."	
RL	Neuron 29:45-55 (2001).	
RN	[4]	TISSUE SPECIFICITY.
RP	Weinmaster G., Roberts V.J., Lemke G.;	
RX	"Notch2: a second mammalian Notch gene."	
RT	Development 116:931-941 (1992).	
RL	[5]	TISSUE SPECIFICITY.
RN	MEDLINE=21331789; PubMed=11438922;	
RX	Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;	
RT	"Expression patterns of Notch1, Notch2, and Notch3 suggest multiple	
RT	functional roles for the Notch-DSL signaling system during brain	
RT	development."	
RL	J. Comp. Neurol. 436:167-181 (2001).	
CC	FUNCTION: Functions as a receptor for membrane-bound ligands	
CC	Jagged1, Jagged2 and Deltal to regulate cell-fate determination.	
CC	Upon ligand activation through the released notch intracellular	
CC	domain (NICD) it forms a transcriptional activator complex with	
CC	RB-J kappa and activates genes of the enhancer of split locus.	
CC	Affects the implementation of differentiation, proliferation and	
CC	apoptotic programs (By similarity). Acts instructively to control	
CC	the cell fate determination of CNS multipotent progenitor cells,	
CC	resulting in astroglial induction and neuron/oligodendrocyte	
CC	suppression.	
CC	SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-	
CC	terminal fragment N(IC) which are probably linked by disulfide	
CC	bonds (By similarity).	

QY 1 GATCTGGGGTGTCCAGGAAAAACCAATCTCGAAGTTAATGGTTTGAAGTATTTT 60
 Db 21 GLRCSQPSGT-----CLNGRCEVANGTEACVSGAFVQRCD----- 59
 QY 61 AAATCCTTGC--TGCC-----GGAGAGCCCC-----GCTCTCCCGGTA-- 98
 Db 60 -PSPCLSTPKNAGTCYVVDHGGIVDYACSPGLGSPGLCLTPLANACLANPCRNNGTCD 118
 QY 99 ----TCAGCGT-----TCCATCTTTGAATCCGGGCT-----CCGC-----GGTC 138
 Db 119 LITLTYKRCPPHSGKSCQAQAPCASN---PCANGGQCLPFESSYICGPPGFHGPCT 175
 QY 139 -TTCCGCTGACAG--CAGCCGAGAGAGCCTGTTTGAATTAAGC----- 182
 Db 176 RQVNECSQNPGLCRHGTCHNEIGSYRC-----ACRATHGPHCELPYVPCSPQCN 229
 QY 183 GGGCTGTG-AAGCCGAGGCGCGCG-----GGGGGGGCGCGAGCGCGGCG--CATTTG 234
 Db 230 GGTCTPTGDTHEACALPFAQONCEENVDCPNCKNGACVDGNTYNCRPPETW 289
 QY 235 -----AATAAGAGCGCTGCTTCCAGGAGGCTCTATAAGTGACCG-----CC 278
 Db 290 QYCTEDVDEQLMPNACQNAQTCHNS-----HGGYNCVGVNGWTGEDCSDNIDDC 339
 QY 279 GCGGC--GAGC-----GTGGCGGCTGTCAGGTCACTGTAGCGACTTCTTTTGGTTTC 331
 Db 340 ASAACTQGATCHDRVASFYCEPHRTG-----LLCHLNACISNPC 381
 QY 332 TTTCTTTTGGGGACCTC-----TGACTACTCCCCAGCATGAGCGCTGAGCCCGGT 387
 Db 382 NEGSNDINPVGKAICTCPRYGTGPAQSQDVDEC-----ALGANPCEHAGKCLNTL 433
 QY 388 G---CGC-GGCTG-----CTACGAGGCGGTGCTGCTGCTGCGAAGCAGTCTGGCCA 437
 Db 434 GSFEQCLQVYTPRCEIDVNECISNPQNDAT-----CLDQIGEFQCLMPGYEYCE 488
 QY 438 --TGCC-----CGGGGCG-----CGAGGAAGGCG--CGCGCAGCTGAGGAGCGC--- 478
 Db 489 INTDECASSPLHNGRCVDMKINEFLCQPKGFSGLCYDVEDCASTPCXKAGKLDGPN 548
 QY 479 ---CTGAGCTTCTGAC-----GACATGAA--CCAC-----TGCT 509
 Db 549 TTTCTV---CTEYVTHCEVIDECDPDPHIGLCKGVATFTCLQPGYTHGHCTNIN 605
 QY 510 AC-----TC-----CGCGCTGGG-----GACTGGTACCGGAGTCC 542
 Db 606 ECHSOPCRHGGTCQDRDNYLCLCLKGTTGPNCEINLDDCASNPDCDSGTCLDKIDGYECA 665
 QY 543 CGAGAGGCACT-----CAG--CTTAGCCAGGTGGAATCTTACAGCGCTC-ATCGACTA 594
 Db 666 CEEGYTGSMCNVNIDECAGSPCHNGTCEGDIAGFT---CRCPEYHDPFCLSEVNECN- 721
 QY 595 CATTCCTGACCTG-----CAGGTAGTCTGCGGAGCCAGCCCTTGACCCCTGATG 647
 Db 722 -SNPCHGACRDLNGYKDCAPWSGTNCD---INNNECESNPVNGGT-----CKDMTS 773
 QY 648 GCGCCACCTTCCCATCCAGACAGCGGAGCTCGCTCCGGAAC--TTGTCACTCTCCACGAC 706
 Db 774 GYVCTCR-----EGFSGFNQNTNNECASNPCLNQGTICIDVAGYKCN-C 817
 QY 707 AAAAGAGCTTTGGCACTGACTGCGCGCTGCTGCTG--ACACTCCAGAAC----- 755
 Db 818 PLPYTGATCEVVLAPCA--TSPCKNSGVCKESEDYSEFSCVPTGQWQQTCEIDINECVKS 876
 QY 756 GCAGGTGC--TGG---CGCCGCTTCTG-----CTGGGACCCCGGGAACCTCTC 799
 Db 877 PCEHGASCONTNGSVYCLCQAGYTGRCNCESIDDCRPNPCHNGGSCIT---DGVNAFCDC 933
 QY 800 CTGCGGGAAGCGGAGCGGAGGGA--TGGGCCCCCACTTGGCCC--TGCCCACTTGACTTC 856
 Db 934 LFGFQGS---AFCEIDINECATNPQNGANCTDCVDSYCTCTPTGFGNHCENNTPDCES 990

QY 857 ACCAAATCC-----CTTC-----CTGGAGACTAAACCTGGT-----GCTCAGG 894
 Db 991 SCFNGGTCVDPGINSFTCLCPGPTGSCYQYVNVNEDSRPCLHGGTCQDSYGYKCTCPOG 1050
 QY 895 -----AGCGAA-----GGACTGTG-----AACTTGTAG----- 917
 Db 1051 YTGLNCQMLVRWCDSAPCKNGGKCWQNTQYHCECRSGMTGFNCVDVLSVCEVAAKRG 1110
 QY 918 -----CCTGAGAGC-----CCTGAGAGC-----CAGACTAGCTCT--GGCCA 945
 Db 1111 DVTLLQHGGLCVDBEDKHCHYCHQAGYTGYSYCEDEVEDCSNPNCGATCTDYLGGFSCK 1170
 QY 946 CCAGCTGGCG--AGCTCAC--CCTGTCTCCACCCACCC----- 981
 Db 1171 CVAGYHSGNSCEINECLSQPCQNGGTCTDLTNTYKCSPPRGTVGVHCEINVDCHPPLD 1230
 QY 982 -----CCAGTCTTAAG--TCTTTTCAGAGCGTGGAGGTGGAGAGGTGGCTGCTC 1033
 Db 1231 PASRSPKCFNNGTCTVDQVGYTCT-----CPPGFVGER-----CEGDVN 1269
 QY 1034 TCCAAACTATG-----C-CAAGCGCG-----CGGCAGAGC--TGGTC---TTC 1070
 Db 1270 ECLSNPCDPRGTQNCVQVNDPFCECRAGHTGRRCEVINGCRKPCRNNGVCAVASNTA 1329
 QY 1071 TGGTCTC-----C-----TTGGAGMAAGTTCTG--TTGCCCTGATTATGACTCT 1115
 Db 1330 RGFTCRPAPFEGATCENDARTCGSLRCLNGGTCISPRSPCTCLUG---SFTGPECQPP 1386
 QY 1116 ATAATAGA-----GTATATA 1130
 Db 1387 ASSPCVGSNPNYNGTCEPTS 1407

RESULT 8
 NTC1_HUMAN
 ID NTC1_HUMAN STANDARD; PRT; 2556 AA.
 AC P48531;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (hml)
 DE (Translocation-associated notch protein TAN-1).
 GN NOTCH1 OR TAN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Mann R.S., Blumweller C.M., Zagouras P.;
 RT "Complete human notch 1 (hml) cDNA sequence.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE OF 1-2444 FROM N.A.
 RX MEDLINE=91347367; PubMed=1831692;
 RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
 RA Smith S.D., Sklar J.;
 RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
 RT chromosomal translocations in T lymphoblastic neoplasms.";
 RL Cell 66:649-661 (1991).
 RN [3]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangi M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794 (1999).
 RN [4]
 RP INTERACTION WITH DTX1.
 RX MEDLINE=98250176; PubMed=9590294;
 RA Matsuno K., Eastman D., Mitsiades T., Quinn A.M., Carcanci M.L.,
 RA Orntlich P., Kadesch T., Artavanis-Tsakonas S.;

[illegible]

FT	DOMAIN	2522	2525	POLY-SER. CLEFTAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).
FT	SITE	1665	1666	
FT	DISULFID	24	37	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	43	57	BY SIMILARITY.
FT	DISULFID	68	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
Query Match 8.9%; Score 674.5; DB 1; Length 2556;				
Best Local Similarity 23.5%; Pred. No. 1.5e-27;				
Matches 374; Conservative 33; Mismatches 633; Indels 549; Gaps 93				
QY	1	GATCTGGG----	GTGCTGCCAGAAAAAGCAAAATCTTGGAAATTAATGTTTTGAGT----	53
DB	28	GETCLNGKCEAAANGTEACVCGGAVGPRCQDPNCL-----	STPCKNAGTCHV 76	
QY	54	-----	GAITTTTAAATCTTCTGCGGAGAGCGCCGCTCT-----	CC 92
DB	77	VDRGVADYACSCALFGSPGLCLPLDNACLINFC-----	RNGTCDLLLTBYKCR 128	
QY	93	CCGGTATCAGCGTTCTCTCATTTCTTTGAATCCGGCGGT-----	CGGC-----	GGTC-TT 140
DB	129	CPEG-----	WGGKSCQADPCASN-----	PCANGGQLPFEASYICHCPSPFHGPTCRD 178
QY	141	CGCGCTGCAGAC--	CAGCGGAGGAAGCCTGTTGCAATTTAAGC-----	GGG 185
DB	179	VNECGKPLCRHGHTCHNEVGSYECV-----	CRAHTGPNCEPVPVPCSPQCGT 232	
QY	186	CTGTG--	AAGCCCAAGCGCGCGG-----	GGGCGGGCCGAGCGGGC--CATTTTGAAT 237
DB	233	CRPTGDTVTHEACLFGFTGQNCENIDPCGNCKNGGACVDGVTNYCFCPEPWGQ-----	290	
QY	238	AAAGAGCGTGCT-----	TCAGGCA-----	GGCTCTATAAGTAGCAGCG-----
DB	291	-----	YCTBDVDECOLMPNACQNGTCHNTHGGYNCVGVNGWTGDCSENIDDC 339	
QY	279	GCGGC--	GAGC-----	GTGCGCGGTTCGAGTCACTGTAGCGGACTTTTGTGTTTC 331
DB	340	ASAACPHGATCHDRVASFYCECPHRTG-----	-----	LLCHLNDACISNFC 381
QY	332	TTTCTTTTGGGCACTTC-----	TGGACTCACTCCCAGCATGAAAGCGCTGAGCCCGGT 387	
DB	382	NEGSNCDTVPVNGKAICTCPSGYTGAPACQDVDEC-----	SLGANPCFHAGKGINFL 433	
QY	388	G--	CGC--GGCTG-----	CTACGAGGCGGTGCTGCTGCTCGGAACGCACTGCGCCA 437
DB	434	GSFECOCLOGYTGPRCEIDVNECVSNPCNDAT-----	CLDQIGFQCVMFGEYSHCE 488	
QY	438	-----	CCGGGCG-----	CGAGGGAAGGC--CGGCGAGCTGAGGAGCG-----
DB	489	VNTDECASPLHNRCGLDKINEFCQECTFTGHLCOYDVCASPTCKNGAKCLDGEN 548		
QY	479	---	CTGAGCTTCTGGAC-----	GACATGAA--CCAC-----
DB	549	TVTCV--	CTEYGTGTHCEVDIDECDPPCHVGSCKDGVATFTCLCRPGYTHHCETIN 605	
QY	508	CTACTC-----	-----	CGGCTGGG-----
DB	606	ECSSQPLRLRGTCQDPDNYALCFCLKGTTGPNCEINLDDCASPSDGSGLDKIDGYECA 665		
QY	543	CGAGAGGCACT-----	CAG--	CTTAGCCAGGTGAAA-----
DB	666	CPGYTGSMCNINIDECAGNPNCHGTTCEDGNGFTCRCEGYHDPCTCLSEVNECNINPC 725		

RX	MEDLINE=98123114; PubMed=9452463;	45	210	DOMAIN	FT
RA	Gao Z., Garbers D.L.;	215	374	DOMAIN	FT
RT	"Species diversity in the structure of zonadhesin, a sperm-specific	377	542	DOMAIN	FT
RT	membrane protein containing multiple cell adhesion molecule-like	547	1170	DOMAIN	FT
RT	domains.";				
RL	J. Biol. Chem. 273:3415-3421(1998).				
RN	[2]	1171	1280	DOMAIN	FT
RP	SEQUENCE OF 4864-5376 FROM N.A.	1281	1669	DOMAIN	FT
RC	TISSUE=Testis;	1670	2056	DOMAIN	FT
RX	MEDLINE=97271566; PubMed=9126492;	2057	2459	DOMAIN	FT
RA	Gao Z., Harumi T., Garbers D.L.;	2460	2579	DOMAIN	FT
RT	"Chromosome localization of the mouse zonadhesin gene and the human	2580	2699	DOMAIN	FT
RT	zonadhesin gene (ZAN).";	2700	2819	DOMAIN	FT
RL	Genomics 41:119-122(1997).	2820	2939	DOMAIN	FT
CC	-!- FUNCTION: Binds in a species-specific manner to the zona pellucida	2940	3059	DOMAIN	FT
CC	of the egg. May be involved in gamete recognition and/or	3060	3179	DOMAIN	FT
CC	signaling.	3180	3299	DOMAIN	FT
CC	-!- SUBUNIT: Probably forms covalent oligomers.	3300	3416	DOMAIN	FT
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein, exclusively on the	3417	3536	DOMAIN	FT
CC	apical region of the sperm head.	3537	3656	DOMAIN	FT
CC	-!- TISSUE SPECIFICITY: Intestis, primarily in haploid spermatids.	3657	3776	DOMAIN	FT
CC	-!- DOMAIN: The MAM domains probably mediate sperm adhesion to the	3777	3892	DOMAIN	FT
CC	zona pellucida.	3893	3992	DOMAIN	FT
CC	-!- DOMAIN: During sperm migration through the reproductive tracts,	4029	4148	DOMAIN	FT
CC	the mucin-like domain might inhibit inappropriate trapping of	4149	4263	DOMAIN	FT
CC	spermatozoa or promoting adhesion to the oviductal isthmus.	4264	4283	DOMAIN	FT
CC	-!- DOMAIN: The WPD domain 2 may mediate covalent oligomerization (by	4384	4503	DOMAIN	FT
CC	similarity to human intestinal mucin MUC2).	4504	4623	DOMAIN	FT
CC	-!- SIMILARITY: Contains 3 MAM domains.	4624	4743	DOMAIN	FT
CC	-!- SIMILARITY: Contains 25 WPD domains.	4744	4863	DOMAIN	FT
CC	-!- SIMILARITY: Contains 1 EGF-like domain.	4864	5261	DOMAIN	FT
CC	-----	5261	5295	DOMAIN	FT
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	5295	5295	DOMAIN	FT
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	5295	5295	DOMAIN	FT
CC	the European Bioinformatics Institute. There are no restrictions on its	5295	5295	DOMAIN	FT
CC	use by non-profit institutions as long as its content is in no way	5295	5295	DOMAIN	FT
CC	modified and this statement is not removed. Usage by and for commercial	5295	5295	DOMAIN	FT
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	5295	5295	DOMAIN	FT
CC	or send an email to license@isb-sib.ch).	5295	5295	DOMAIN	FT
CC	-----	5295	5295	DOMAIN	FT
DR	EMBL; U97068; AAC26680.1; "	5295	5295	DOMAIN	FT
DR	EMBL; U8190; AAC53125.1; "	5295	5295	DOMAIN	FT
DR	PIR; T42215; T42215.	5295	5295	DOMAIN	FT
DR	MGI; MGI:106656; Zan.	5295	5295	DOMAIN	FT
DR	InterPro; IPR006209; EGF like.	5295	5295	DOMAIN	FT
DR	InterPro; IPR003645; FOLN.	5295	5295	DOMAIN	FT
DR	InterPro; IPR006210; IEGF.	5295	5295	DOMAIN	FT
DR	InterPro; IPR000998; MAM domain.	5295	5295	DOMAIN	FT
DR	InterPro; IPR002919; TIL_Cysrich.	5295	5295	DOMAIN	FT
DR	InterPro; IPR003328; TIL_Cysrich.	5295	5295	DOMAIN	FT
DR	InterPro; IPR001007; WVF_C.	5295	5295	DOMAIN	FT
DR	InterPro; IPR001846; WVF_D.	5295	5295	DOMAIN	FT
DR	Pfam; PF00629; MAM; 3.	5295	5295	DOMAIN	FT
DR	Pfam; PF01826; TIL; 25.	5295	5295	DOMAIN	FT
DR	Pfam; PF02345; TIL; 25.	5295	5295	DOMAIN	FT
DR	Pfam; PF00094; vwd; 4.	5295	5295	DOMAIN	FT
DR	SMART; SM00181; EGF; 2.	5295	5295	DOMAIN	FT
DR	SMART; SM00274; FOLN; 11.	5295	5295	DOMAIN	FT
DR	SMART; SM00137; MAM; 2.	5295	5295	DOMAIN	FT
DR	SMART; SM00214; VWC; 17.	5295	5295	DOMAIN	FT
DR	SMART; SM00216; WVD; 4.	5295	5295	DOMAIN	FT
DR	PROSITE; PS00022; EGF 1; 1.	5295	5295	DOMAIN	FT
DR	PROSITE; PS01186; EGF 2; 18.	5295	5295	DOMAIN	FT
DR	PROSITE; PS00026; EGF 3; 1.	5295	5295	DOMAIN	FT
DR	PROSITE; PS00740; MAM_1; FALSE_NEG.	5295	5295	DOMAIN	FT
DR	PROSITE; PS00060; MAM_2; 3.	5295	5295	DOMAIN	FT
KW	Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;	5295	5295	DOMAIN	FT
KW	Repeat.	5295	5295	DOMAIN	FT
FT	SIGNAL.	1	17	POTENTIAL.	FT
FT	CHAIN.	18	5376	ZONADHESIN.	FT
FT	DOMAIN.	18	5310	EXTRACELLULAR (POTENTIAL).	FT
FT	TRANSMEM.	5311	5337	POTENTIAL.	FT
FT	DOMAIN	5338	5376	CYTOPLASMIC (POTENTIAL).	FT

Query Match 8.9%; Score 670.5; DB 1; Length 5376;

Best Local Similarity 20.1%; Pred. No. 46-27; Indels 883; Gaps 106;

Matches 395; Conservative 77; Mismatches 612; Indels 883; Gaps 106;

QY 11 GCT-GC-CAGGAAAAA--GCAAATCTGGAAGTTAATGTTTGTAGTGATT----- 57

RN RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
 Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
 RT deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.
 RL Nature 393:537-544 (1998).
 [2]
 RN RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains".
 RT J. Bacteriol. 184:5479-5490 (2002).
 [3]
 RN RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis".
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 RL -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGFS
 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z80225; CAB02341.1; --
 DR EMBL; AE007103; AAK47026.1; ALT_INIT.
 DR EMBL; BX248343; CAD94852.1; --
 DR PIR; F70963; F70963.
 DR TIGR; MT2712; --
 DR TubercuList; Rv2634c; --
 DR InterPro; IPR000084; PE_region.
 DR Pfam; PF00934; PE; 1.
 DR ProDom; PD001223; PE_region; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 51 51 V -> L (IN REF. 2).
 FT CONFLICT 63 63 Q -> H (IN REF. 2).
 FT CONFLICT 274 274 A -> T (IN REF. 2).
 FT CONFLICT 274 274 A -> T (IN REF. 2).
 SQ SEQUENCE 778 AA; 63131 MW; DAB20F5E8E4999E7 CRC64;
 Query Match 8.9%; Score 668; DB 1; Length 778;
 Best Local Similarity 29.3%; Pred. No. 1.3e-27;
 Matches 295; Conservative 12; Mismatches 317; Indels 384; Gaps 44;
 QY 133 CGCGTCTCCGCGTCACACCGAGGAGCGCTGTTTCAATTTAAGCGGCTGTGAA 192
 DB 133 GPGGLLLGNGNG---GSGAPGQPGAGGAGL-----IGNG--GTGCK 171
 QY 193 CGCCACAGCGCGCGGGGG--CGGGCCGAGCGGGCC--ATTTGAATAAGAGCGGTG 248

Db 172 GGDGLVSGAAGGVGGRGGWLLNGGTTGAGGAAGATLVGGTGGVGGATGLIGSGGFGG-- 230
 QY 249 CTTCCAGGAGGCTCTATAAGTGACCGCGCGGAGCGCTGCGCGTTGAGGTCAC 308
 Db 231 -----AGGAAGAGVTGGVGGSGGVGFGNGGFG-----GAG-GLGAAGGVGAAYF 278
 QY 309 GTAGCGGACTCTTTTGGTTTCTTCTTTCTTTGGGCACTCTGGACTCACTCCACGCA 368
 Db 279 GTGGGG-----VGG----- 288
 QY 369 TGAAGCGCTGAGCCCGGTGCGCGGTGTCTACAGCGGTGTCTGCTCTCGGAACGCA 428
 Db 289 DGAPGGDG--GAGPLLING-GVGGLLGAG--AGNGGAG-----GMLLG--DGGAGGOG 337
 QY 429 GTCTGGCCATCGCCCGGGCGGAGGAGGAGGCGCGGAGCTGAGGAGCGCTGAGCTTGC 488
 Db 338 GPAVAGVL-----GMPGAGN-----GGNANWFGSGGAGGQGGTLAGTNGV 380
 QY 489 TGAAGACATGAACCACTGCTACTCCCGCTGCGGGAACCTGGTACCGGAGTCCCGAGAG 548
 Db 381 NPGSIANPNTGANGTDSG-----NGNQTGGNGGPGAGGVG--EAGGVG-----GQG 426
 QY 549 GCACCTAGCTTAGCCAGGTGGAATCTACAGCGGTCTACGACTCATCTCGACCTGC 608
 Db 427 GLGESLDG-----NDGTGG-----KGG 443
 QY 609 AGGTAGTCTTGGCGGAGCGCCCTGAGCCCTGATGCGCCCGCCCACTTCCCATCCAGA 668
 Db 444 AGTAGT--DGAAGAG--GAGGIGETDGSAGVATGEGG-----DGA 483
 QY 669 CAGCGGAGCTCGCTCCCGAACTTGTCTATCTCCAAAGCAAAAGAGCTTTTGCACGAC 728
 Db 484 TGGVGG-----VGG-----AGKGGQGHNTGVG-DAFGD- 513
 QY 729 TCGGCGGTGTCTGACACCTCCAGAACGAGGTGTCTGCGCGCCCTTCTGCTGGGACCCC 788
 Db 514 --GGIGDG-----NGALGAGGNGGTGAG-----GNGGRGGLI 547
 QY 789 GGGAACTCTCTCCCGAAGCGGAGGAGATGGGCCCCCACTTGCCTCCCTGCCAC 848
 Db 548 GNGCA-----GGAGTGTGGGGAAGFAGG--VGG----- 575
 QY 849 TTGACTTCAAAATCCCTCTCTGAGACTAAACCTGTCTCAGGAGCGAAGACTGTG 908
 Db 576 -----AGBELTDAAGTAEGTG-GLGGLGVGGTGMGSGG 612
 QY 909 AACTTGTAGCTGAAGAGCCAGACTAGCTCTGCGCCACCACTGGCGACGTCACTCCTGC 968
 Db 613 VGGNGAAGSLILGLGGGGAGVGTTGGTGGI--AGNGG----- 652
 QY 969 TCCACCCCGCCCAAGTCTAAGTCTTTTTCAGAGCGTGGAGGT--GTGGAAGAGTGG 1027
 Db 653 -----AGGAGTTT---GGGATGGGGGTGGVGGAGTGGTGG 686
 QY 1028 CTGCTCTCCAAACTATGCCAAGCGCGGCGAG-AGCTGGTCTTCTGCTCTCTTGGAGAA 1086
 Db 687 AGTT-----GSGGAGGLIGWAGAGGTGAGGTGG-----QGGLG-G 723
 QY 1087 AGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1134
 Db 724 QGNGGNGGTG-----ATCGCGDFALGNGGAGGAGGSGGSGG 762
 RESULT 11
 LORI_MOUSE
 ID LORI_MOUSE STANDARD; PRT; 481 AA.
 AC P18165;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Loricrin.
 GN LOR.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90275605; PubMed=2190691;
 RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
 RA Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinart P.M.,
 RA Yuspa S.H., Roop D.R.;
 RT "Identification of a major keratinocyte cell envelope protein,
 RT loricrin,"
 RL Cell 61:1103-1112 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/C;
 RX MEDLINE=95256248; PubMed=7738016;
 RA DiSepio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
 RA Roop D.R.;
 RT "The proximal promoter of the mouse loricrin gene contains a
 RT functional AP-1 element and directs keratinocyte-specific but not
 RT differentiation-specific expression,"
 RL J. Biol. Chem. 270:10792-10799 (1995).
 CC -!- FUNCTION: Major keratinocyte cell envelope protein.
 CC -!- SUBUNIT: Monomers are crosslinked by disulfide and N-(gamma-
 CC glutamyl) lysine isodipeptide bonds.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M34398; AAA39444.1; -;
 DR EMBL; U09189; AAA82152.1; -;
 DR PIR; A35628; A35628.
 DR HSP; P02876; 9WGA.
 DR MGD; MGI:96816; LOR.
 KW Keratinization.
 SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;
 Query Match 8.8%; Score 666; DB 1; Length 481;
 Best Local Similarity 29.9%; Pred. No. 1-2e-27;
 Matches 206; Conservative 48; Mismatches 190; Indels 244; Gaps 35;
 QY 147 CAGACACCCGAGGAGCCTGTTTCAATTAAGCGGCGTGTGAACGCCACCGCGCGG- 205
 DB 11 CPPVCGKTSGGGGGGGGYSG-----GGSGCGGSSGGSSCG--GGGGGSGGG 59
 QY 206 --CGGGGCGGG--GCGGAGGCGGCCCATTTTGAATAAGAGCGGTGCTTCCAGGCA 259
 DB 60 SSCGGGGSGGVKYSGGGGSSCGG-----YSGGGG-----SSCGGYS 101
 QY 260 GCCTCTAATAGTACCGCGCGCGGCGGCGTGTGCGGTGACGTGCTGACGCGACTT 319
 DB 102 GG-----GGSSCG-----GGYSGGGGSSCGGSGSGGGSSCGGGSGGG-- 143
 QY 320 CTTTGTGTTTCTTCTTCTTGTGGGCACTCTGCACTCCACGAGCTGAAGCGGTG 379
 DB 144 -VKYSGG-----GGGGSSC-----GG-----GSSGGGGSSCGG 173
 QY 380 ACCCGGTGCGCGCTGCTACGAGCGCGTGTGCTGCTGCTGCGAAGCAGTCTGCGCATC 439
 DB 174 GS-----GGGGYCGGSSG--GGSSGCGCGGSGG--GKYSGGGGSSCG--GGYSGG----- 219
 QY 440 GCCCGGGGCGAGGAGGCGCGCGGCGAGCTGAGGAGCGGTGCTGCTGCGACGACATG 499
 DB 220 GSSGSSGSC--GGYSGGGSSCGGGGSGGGGSGGGTSCGG-----GSSGGGGGSSG 269
 QY 500 AACCACTGCTACTCCGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 559
 DB 270 YCCGSGG-----GSSGGSSCGGYS-----GGGGSSCGGYS-----GSSGGSSCGG 311

QY 560 AGCCAGGTGGAATCTCTACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 619
 DB 312 -----GSSGGG-----SSCGSGGG-----GYSGG-----GG 333
 QY 620 GCGAGCCAGCCCTGACCCCTGATGCCCCACCTTCCATCCAGACAGCCGAGCTC 679
 DB 334 GSCGGG-----SSGGGGYSSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQ 375
 QY 680 GCTCCCGGAACTGTCATCTCTCAACGACAAAGGAGCTTTTGCCTGCTGCTGCTGCTG 739
 DB 376 GGGSSG-----GGGG-----GYSGGGGSSGGCGG----- 403
 QY 740 CTGACACCTCCAGAGCGAGTGTGCGCGCCCGTCTTCCCTGCGGACCCCGGAACTCTC 799
 DB 404 -----YSGGGGCGGGS--SGSGGGCGGSSG--GSGGG-----CGGYS----- 440
 QY 800 CTGCCGGAAGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 827
 DB 441 -----GGGG--GSSCGGSSGGSGGGG 461

RESULT 12

NTC1 MOUSE
 ID NTC1 MOUSE STANDARD; PRT: 2531 AA.
 AC Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9ROX7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A)
 GN (nt14) (p300).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch,"
 RL Genomics 15:259-264 (1993).
 RN [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RX MEDLINE=93050801; PubMed=1426644;
 RA Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
 RT "Expression analysis of a Notch homologue in the mouse embryo,"
 RL Dev. Biol. 154:377-387 (1992).
 RN [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development,"
 RL Development 115:737-744 (1992).
 RN [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues,"
 RL Exp. Cell Res. 204:364-372 (1993).
 RN [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE=99364499; PubMed=10437788;

RA Lee J.S., Ishimoto A., Yanagawa S.I.;
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
 RT to induction of HES-1 in a mouse T lymphoma cell line, DU-3.";
 RL FEBS Lett. 455:276-280(1999).
 CC
 RA MEDLINE=98029496; PubMed=9384671;
 RX MEDLINE=98029496; PubMed=9384671;
 RA Messerle M., Folio M., Nehls M., Eggert H., Boehm T.;
 RT "Dynamic changes in gene expression during in vitro differentiation of
 RT mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
 CC
 RA [7]
 RX SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
 RT MUTAGENESIS OF 1651-ARG--ARG-1654.
 RA MEDLINE=98118619; PubMed=9653148;
 RX MEDLINE=98118619; PubMed=9653148;
 RA Logeat F., Bessia C., Brou C., Lebaill O., Jarriault S., Seidah N.G.,
 RA Israel A.;
 RT "The Notch1 receptor is cleaved constitutively by a furin-like
 RT convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
 CC
 RA [8]
 RX PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RA MEDLINE=21523956; PubMed=11518718;
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 CC
 RA [9]
 RX POST-TRANSLATIONAL PROCESSING.
 RA MEDLINE=2137376; PubMed=11459941;
 RX MEDLINE=2137376; PubMed=11459941;
 RA Mizutani T., Taniuchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC
 RA [10]
 RX INTERACTION WITH DTX1 AND DTX2.
 RA MEDLINE=21123790; PubMed=11226752;
 RX MEDLINE=21123790; PubMed=11226752;
 RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
 RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Teakonas S.,
 RA Okano H., Matsuno K.;
 RT "Murine homologs of Deltex define a novel gene family involved in
 RT vertebrate Notch signaling and neurogenesis.";
 RL Int. J. Dev. Neurosci. 19:21-35(2001).
 CC
 RA -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation. May be involved in mesoderm
 CC development, somite formation and neurogenesis. Involved in the
 CC maturation of both CD4+ and CD8+ cells in the thymus.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(BC) which are probably linked by disulfide
 CC bonds. Interacts with DTX1 and DTX2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q01705-1; Sequences=Displayed;
 CC Name=2;
 CC IsoId=Q01705-2; Sequences=VSP_001402, VSP_001403, VSP_001404;
 CC Notes=Experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
 CC thymus. Expressed at lower levels in the spleen, bone-marrow,
 CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
 CC muscle, kidney and heart.
 CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
 CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
 CC endothelial cells, while much lower levels are seen in the

CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
 CC the neuroepithelium. At 13.5 dpc expressed in the surface
 CC ectoderm, eye and developing whisker follicles.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(BC). Following
 CC ligand binding, it is cleaved by TNP-alpha converting enzyme called
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 36 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 Ank repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z11886; CAA77941.1; -
 CC EMBL; I02613; AAK14898.1; -
 CC EMBL; X68278; CAA48339.1; -
 CC EMBL; AJ238029; CAB40733.1; -
 CC EMBL; X82562; CAA57909.1; -
 CC EIR; A46019; A46019; -
 CC EIR; B49175; B49175; -
 CC HSSP; P00740; IEDM.
 CC
 CC MGD; MGI:97363; Notch1.
 CC GO; GO:0005887; C:integral to plasma membrane; IC.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC GO; GO:0030154; P:cell differentiation; IMP.
 CC GO; GO:0007386; P:compartment specification; IMP.
 CC GO; GO:0007219; P:N signaling pathway; IC.
 CC GO; GO:0045944; P:positive regulation of transcription from P. . . ; IDA.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx hydroxy_1_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 7.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGF_BLOOD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 24.
 CC SMART; SM00004; NL; 2.
 CC PROSITE; PS50297; ANK_REPEAT_1.
 CC PROSITE; PS50088; ANK_REPEAT_2.
 CC PROSITE; PS00010; ASX_HYDROXYL; 22.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 27.
 CC PROSITE; PS50026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 21.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing.

FT	SIGNAL	1	18	POTENTIAL.
FT	CHAIN	19	2531	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT	CHAIN	1711	2531	NOTCH EXTRACELLULAR TRUNCATION.
FT	CHAIN	1744	2531	NOTCH INTRACELLULAR DOMAIN.
FT	CHAIN	1744	2531	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	19	1725	EXTRACELLULAR (POTENTIAL).
Query Match				
Best Local Similarity 8.8%; Score 661.5; DB 1; Length 2531;				
Matches 352; Conservative 36; Mismatches 596; Indels 549; Gaps 85;				
QY	1	GATCTGGGTGCTGCGAGAAAGCAAAATTCCTGGAAGTAATGTTTGGTGGTATTTT	60	
DB	21	GLRCSFSGT-----CLNGRCEVASTEACVAGSGFVGRCODPN-----	61	
QY	61	AAATCCTTGC--TGGCGGAGAGCCCGCTCTCCCGGTATCAGCGCTTCCTCATCTTT	118	
DB	62	---PCLSTRCKNAGTCVVDHGGVNDYACSLPFGSP-----LCLTFLDKP	105	
QY	119	GAATCCGCGCTC-----CGCGTCTTCGGCGTCAGACAGCGGAGGAAGCCT	167	
DB	106	CLANPCRNGG--TCDLLTLEYKRCRSPGWSGKSCQADPCASNPCAN-----	151	
QY	168	GTTTGCAATTAAGCGGCGTGTGAAGCCCA-----GGGC-----CGGCGGGG--	218	
DB	152	-----GGQCLPFESSYICRPPGFHGTCTCRQDVNECSQNFGLCRHGGCH	196	
QY	219	---GAGCGGCGCATTTTGAATAAGAGCGGTGC---CTTCAGGAGGCTCTATAAGTG	272	
DB	197	NEIGSYRC--ACCATHT-----GPHCELFYVPCSPQNGATCRPTGDTTH	241	
QY	273	ACGCGCGGCGAGCGTGGCGGCTTGCAAGTCACTAGCGGACTTCTTT-----	323	
DB	242	EC--ACLPFAGQCEB-----NVDDCPGNCK-----NGACVDGVNTYNCRCPEV	287	
QY	324	TGTTT-----TCTTCTTTTGGG--GCACCT--CTGACCTCACTCCCGCAGTG	370	
DB	288	TGYCTEDVDEQLMPNACNAGTCHNTHGGYNCVNVGTGDECSENIDDCASACFOG	347	
QY	371	AAGCGCTGAGCCC--GTCGGCGCTGTCTA--CGAGCGGTGTGCTGCTGCTGCG--	425	
DB	348	ATCHDRVASFYCEPHGRTGLLHLKACISNFCNEG-----SNCNTNVNKRIC	398	
QY	426	CGAGCTTGCCCA--TCGCGCGGCGCCGAGGAGAGGCGCGGAGGAGCC--GCTGA	482	
DB	399	TCPSGTGPACQDVDECDLGNRCEHA-----GKCLNTLGSFECQCLQGTGP	447	
QY	483	GCTTGTGAGCAGCATGAC-----CAC-----TGCTACTC-----C--	514	
DB	448	GCEIDVNECISNPNQNDATCLDQGEFQCICMPGYEYVCEINTECASSPCLNHHOMD	507	
QY	515	-----CGCTTGGCG-----GAACTGTGTACCG--GAGTCCCGAGAGGAC-----	552	
DB	508	KIHEFPQCQPKGFNGHLQYDVBECASPTCKNGAKCLDGNVTVCVCTEGYGTGTCHEVDI	567	
QY	553	-----TC-----ACITAGCCAGGTGAATCTCTACAGCC-----GTCA-----	589	
DB	568	DECDPDPCHYSGCKDGVATFTCLQPGYTHHCETNINECHSQPCRHHGTCQDRDINSYLC	627	
QY	590	GACTACATCTCGACCTCCAGG--TAGTTC-----CTGGCGGAGC-----CAGCC	631	
DB	628	-LCLKGTTPNCEINLDCASNPCDSGTLKIDYECACEBPYGTGSMCNVNDICAGSP	686	
QY	632	CCTGAGACCCCTGATGGGCC-----CCACTTCCCATCCACAGA--GCCGAGCTC--	679	
DB	687	CHNGTCTEDGIAGFT--CRCPEGYHDPFTCLSEVNECSNPNCHGACRDGLNGYKCDAPG	744	
QY	680	--GCTC-----CGGAATTT--GTCA-----TCTC-----CAAGC	704	
DB	745	WSGTNCDINNECESNPVNGTCKDVTSGYVCTCRGFGSPNQTNINECASPNCLNOG	804	
QY	705	ACAAAGAGACT-----TTTGCACTGACTCGCGCGTGTCTCTG-----ACACTCCAGA	753	
DB	805	TCIDDVAGYKNCPLPYTGATCEVVLAFPCATSPCKNSGVCKESEDYFSFCVCTGMOGQ	864	

RESULT 13

NTC3	HUMAN	STANDARD;	PRT;	2321	AA.
ID	NTC3_HUMAN	Q9UEB3; Q9UEB3; Q9UEB3; Q9UEB3;			
AC	Q9UEB3; Q9UEB3; Q9UEB3; Q9UEB3;				
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Neurogenic locus notch homolog protein 3 precursor (Notch 3).				
GN	NOTCH3.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97032728; PubMed=8878478;				
RA	Joutei A.; Corpechot C.; Ducros A.; Vahedi K.; Chabrier H.; Mouton P.;				
RA	Alanowitch S.; Domenga V.; Cecillon M.; Marechal E.; Maciazek J.;				
RA	Vaysiere C.; Cruaud C.; Cabanis E.-A.; Ruchoux M.M.; Weissenbach J.;				
RA	Bach J.-F.; Bousset M.-G.; Tournier-Lasserre E.;				
RT	"Notch3 mutations in CADASIL, a hereditary adult-onset condition				
RT	causing stroke and dementia.";				
RL	Nature 383:707-710 (1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Gunel M.; Artavanis-Tsakonas S.;				
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Lamerdin J.E.; McCreedy P.M.; Skowronski E.; Adamson A.W.;				
RA	Burkhart-Schultz K.; Gordon L.; Kyle A.; Ramirez M.; Stillwagen S.;				

Phan H., Velasco N., Garnes J., Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T., Tranheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.,
 RT Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in 19p13.1.";
 RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RL [4]
 RN VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141; ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212; GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728; CYS-985; CYS-1006; CYS-1031; CYS-1031 AND ARG-1261, AND VARIANTS ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
 RX MEDLINE=9049733; PubMed=9388399;
 RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H., Vayssiere C., Cruaud C., Maciazek J., Weissbach J., Bousser M.-G., Bach J.-F., Tournier-Lasserre E.;
 RA "Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients.";
 RT Lancet 350:1511-1515 (1997).
 RL [5]
 RN VARIANT CADASIL 114-GLY--PRO-120 DEL.
 RP MEDLINE=20264473; PubMed=10802807;
 RX Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C., Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.;
 RA "Splice site mutation causing a seven amino acid Notch3 in-frame deletion in CADASIL.";
 RT Neurology 54:1874-1875 (2000).
 RL [6]
 RN IDENTIFICATION OF LIGANDS.
 RP MEDLINE=99180765; PubMed=10079256;
 RX Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L., Banks A., Leiman J., Ward D., Ieh-Horowitz D., Artavanis-Tsakonas S.;
 RA "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794 (1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytic processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult tissues.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(EC). Following ligand binding it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NtCD) from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal dominant arteriopathy with subcortical infarcts and stroke and leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type of stroke and dementia of which key features include recurrent subcortical ischemic events and vascular dementia.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 34 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U97669; AAB91371.1; -
 DR EMBL; AF058900; AAC14346.1; -
 DR EMBL; AF058881; AAC14346.1; JOINED.
 DR EMBL; AF058882; AAC14346.1; JOINED.
 DR EMBL; AF058883; AAC14346.1; JOINED.
 DR EMBL; AF058884; AAC14346.1; JOINED.
 DR EMBL; AF058885; AAC14346.1; JOINED.
 DR EMBL; AF058886; AAC14346.1; JOINED.
 DR EMBL; AF058887; AAC14346.1; JOINED.
 DR EMBL; AF058888; AAC14346.1; JOINED.
 DR EMBL; AF058889; AAC14346.1; JOINED.
 DR EMBL; AF058890; AAC14346.1; JOINED.
 DR EMBL; AF058891; AAC14346.1; JOINED.
 DR EMBL; AF058892; AAC14346.1; JOINED.
 DR EMBL; AF058893; AAC14346.1; JOINED.
 DR EMBL; AF058894; AAC14346.1; JOINED.
 DR EMBL; AF058895; AAC14346.1; JOINED.
 DR EMBL; AF058896; AAC14346.1; JOINED.
 DR EMBL; AF058897; AAC14346.1; JOINED.
 DR EMBL; AF058898; AAC14346.1; JOINED.
 DR EMBL; AF058899; AAC14346.1; JOINED.
 DR EMBL; AC004257; AAC04897.1; -
 DR EMBL; AC004663; AAC15789.1; ALT_INIT.
 DR PIR; S78549; S78549.
 DR HSSP; P00740; IEDM.
 DR Genew; HGNC:7883; NOTCH3.
 DR MIM; 600276; -
 DR MIM; 125310; -
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx hydroxyl_s.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR008800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 19.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 25.
 DR PROSITE; PS00026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_CA; 16.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Disease mutation.
 KW SIGNAL 1 39
 FT CHAIN 40 2321
 FT CHAIN 1629 2321
 FT CHAIN 1662 2321
 FT CHAIN 40 1643
 FT DOMAIN 40 1643
 FT POTENTIAL.
 FT NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 FT NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT EXTRACELLULAR (POTENTIAL).
 FT

FT	TRANSMEM	1644	1664	POTENTIAL.	
FT	DOMAIN	1665	2321	CYTOPLASMIC (POTENTIAL).	
FT	DOMAIN	40	77	EGF-LIKE 1.	
FT	DOMAIN	78	118	EGF-LIKE 2.	
FT	DOMAIN	119	156	EGF-LIKE 3.	
FT	DOMAIN	158	195	EGF-LIKE 4.	
FT	DOMAIN	197	234	EGF-LIKE 5.	
FT	DOMAIN	236	272	EGF-LIKE 6.	
FT	DOMAIN	274	312	EGF-LIKE 7.	
FT	DOMAIN	314	350	EGF-LIKE 8.	
FT	DOMAIN	351	389	EGF-LIKE 9.	
FT	DOMAIN	391	429	EGF-LIKE 10.	
FT	DOMAIN	431	467	EGF-LIKE 11.	
FT	DOMAIN	469	505	EGF-LIKE 12.	
FT	DOMAIN	507	543	EGF-LIKE 13.	
FT	DOMAIN	545	580	EGF-LIKE 14.	
FT	DOMAIN	582	618	EGF-LIKE 15.	
FT	DOMAIN	620	655	EGF-LIKE 16.	
FT	DOMAIN	657	693	EGF-LIKE 17.	
FT	DOMAIN	695	730	EGF-LIKE 18.	
FT	DOMAIN	734	770	EGF-LIKE 19.	
FT	DOMAIN	771	808	EGF-LIKE 20.	
FT	DOMAIN	810	847	EGF-LIKE 21.	
FT	DOMAIN	849	885	EGF-LIKE 22.	
FT	DOMAIN	887	922	EGF-LIKE 23.	
Query Match					8.7%; Score 656; DB 1; Length 2321;
Best Local Similarity					23.7%; Pred. No. 1.2e-26;
Matches 344; Conservative 40; Mismatches 566; Indels 502; Gaps 80;					
QY	158	GAGAAAGCCTTTGCAATTAACGGGGCTG-TGAACGCCAGCGCGGGGGGGGGG	216		
DB	35	GPAAAPPLDGP-CAN-----GGRCQLPSREAAACLPFGWVERQLEDPCHSGP	86		
QY	217	CCGAGGGGGCCATTTTGAATAAGAGCGTGCCTCCAGGACAGGCTCTATAGTACCG	276		
DB	87	CAGRGVQSSVAGTARFSCRCPRGPDCLSPDCLSPFCAGARCSVGPDRFLC--	144		
QY	277	CCGCGGGAGCGTGC-CGCGTTCCAGTCACTAGCGGACTCTTTTGGTTTCTTTC	335		
DB	145	SCPFGYQGRSCRSDVDECRVGEPCRHGTCCLNT-----PGRFCQCPAGYTGPLCENPA	198		
QY	336	TCTTTGGGCACTTC-TGGACTCACTCCAGCATGAAG-GC-----GCTGAGCCCGGT-	387		
DB	199	V-----PCAPSPRNGGTCRQSGDLTYDCACLPFGEGQCNVNVDDCFGHRCLNGT	251		
QY	388	-----CCGC-GGCTG--CT-----ACGAGCGGTGTGCTGCTGCGGAAG	426		
DB	252	VDGVNTYNCQPPPEWTGQCTEDVDECOLQPNACHG-----GTCFNTLGHSCV	301		
QY	427	CAGTCTGCCA-----TCGCGCGGGC-----CGAGGGAAG-----GGCGCGGC-	465		
DB	302	CVNGWTGSCSNIDDDCATVCFRGATCHDRVASFYCACPMGKTGLLCHLDDACVSNPCH	361		
QY	466	--AGCTGAGGAGCCGCTGAGCTGTGTGAC-----GACATGAACACTG-C--TACTCCG	516		
DB	362	EDAICDTNPNVNGRAICT--CPPGFTGACDQDDVDECSIGANPCHEHLGRVNTQGSFLCQ	418		
QY	517	CCTGCGGAACTGTGTACCG-----GAGTCCCGAGAGS-C-----ACTCAGC	557		
DB	419	CGRGVTPRCEFDVNECLSGPCRNQATCLDRIGQPTCICMAGFTGTVCEDVIDECQSSPC	478		
QY	558	TTAGCCAGGTGAAATCTCATAGCGC-----GTCACTGACTACATTCCTGACCTG-----	607		
DB	479	VNGGVCKVRVNGFSCTCFSGFGSTCQLDVDECA-----STPCRNGAKCVDPQDGYE	530		
QY	608	--CAGTAGTCTCT-----GGCGGACGAGCCCT-----	634		
DB	531	CRCAEGFGLCDRNVDSCPPCHHRCVDPGASFCACAPGYGTRCSEQVDECRSQP	590		
QY	635	-----GGAC-----CCCTGATGCGC-----CCACCT-----TCCCATCCA	666		
DB	591	CRHGKCLDLVDKYLRCPSGTTGNCVNEWIDDCASNPTCTFGVCRDGINRYDCVCQPGFT	650		

QY	667	G-----ACAGC-CGAGCTC-----CCTC-----CGAACTTGTC- 694		
DB	651	GPLCNVEINCASSPCGEGSCVDGNGFRCLCPGSLPLCLPPSPHPCAEPCSHGICY 710		
QY	695	-----ATCTC-----CAACACAAAAGGAGCTTTTGC-----CA 723		
DB	711	DAPGFRVCBPGWSPRCSQSLARDACESQPCRAGTCSDDGNGFFHCTCPPGVQGRQCE 770		
QY	724	CTGACTCGGCGTGTCT-CTGACACCTC-----CAG-AAAG-----CA 758		
DB	771	LLSPCTNPCEHGRCHESAPGQLPVCSPQGWGRQDQVDECAAGPAPCGPHGICNTLA 830		
QY	759	GGTGTCT--GGCGCCCGTTTGTCTG-----CTGGAGCCCGGAACTCTCTCTG-----CCGG 806		
DB	831	GSFSTCHGGYTGSPSCDQINDCDPNFLNGS---CODGVGSPSCCLPFGAFAGCARD 887		
QY	807	AAGCGGACG-----GCAGGATGG-GC-----CCCAACTTCGCCCTGC--C 845		
DB	888	VDELSNPFGGTCTDHVASTCTCPGYGFHCEQDLPCDSPSSCFNGGTCVDGVNFS 947		
QY	846	CACTTGACTCACCAATCCCTTCTGAGACTAACTGGTGCTCA---GGAGCGAA-- 900		
DB	948	CLCRPG-YTGAHCQHEADPCLSRPLHG-GVCSAA---HFGFRCTCLSFPGQCTLVD 1002		
QY	901	-----GGACTGTGAACCTGTAGCTGAAGAC-----CAGAGCTAG-----CTC 939		
DB	1003	WCSROPQONGRCVQTGAYCL-----CPGWSGRCLDIRSLPCREAAAIQVRLQQLQA 1057		
QY	940	TGGC-----CACCAGCTGGC-----GAC-----GTCACCTGCTCCC----- 972		
DB	1058	GSQCVDEDSHYVCPEORTGSHCEQVDPCLAPQCQHGTCRGYMGGMCECLPGYNGD 1117		
QY	973	-----ACCCAC-----CCCCAAGTCTTAAGTCTTTTTCAGAGCGTGGAG 1012		
DB	1118	NCEDDVDCAQPCQOHGSCIDLVARYLCSPPGT---LGLVLEINE---DDGPGPPL 1170		
QY	1013	GTGTGGAAGGAGTGGCTCTCTCCAAACTATGC-----CAAGGC-----GG 1053		
DB	1171	DSGPCLHNGTCVDLVGGFRCT-CPGYTGLRCEADINECRSGACHAAHTRDCLDPPGG 1229		
QY	1054	CGCGAGACTGTCTT---CTGGTC-----TCTTGGAGAAAGTCTCTGTGTC--- 1100		
DB	1230	FRCLCHAGFSGRPCVTLSPCESQPCQHGQCRPSPGF---GGTFT---CHCAQPFW 1282		
QY	1101	-----TGATTTTGAATC-----TATATAGAGTATATA- 1130		
DB	1283	GPRCERVARSCELCQPVPCQPTPRGPRCACPGLSGPSCRFPSPGSPGASNASCAA 1342		
QY	1131	-----GGTTTG-----TACCTTTTACAGGAGGTGACTTTCTGTAA----- 1169		
DB	1343	PCLHGGSCRPAPLAPFFRCACAGWTGPRCEAPAAPEVSEPRCPRAACQAKRGDQRCD 1402		
QY	1170	--CAATGCGATG 1179		
DB	1403	RECNSPGCGWDG 1414		
RESULT 14				
BAR3 CHITE				
ID	BAR3	CHITE	STANDARD;	PRT; 1700 AA.
AC	Q03376;			
DT	01-OCT-1993	(Rel. 27, Created)		
DT	01-OCT-1993	(Rel. 27, Last sequence update)		
DT	15-MAR-2004	(Rel. 43, Last annotation update)		
DE	Barbani ring protein 3 precursor.			
GN	BR3.			
OS	Chironomus tentans (Widge).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Nematocera; Chironomidae;			
OC	Chironomidae; Chironominae; Chironomus.			
OX	NCBI_TaxID=7153;			
RN	[1]			

SEQUENCE FROM N.A.
 RP TISSUE=Salivary gland;
 RX MEDLINE=90172404; PubMed=1689777;
 RA Paulsson G.; Lendahl U.; Galli J.; Ericsson C.; Wieslander L.;
 RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
 repetitive structure split by many introns.";
 RL J. Mol. Biol. 211:331-349(1990).
 CC -!- FUNCTION: Used by the larvae to construct a supramolecular
 structure, the larval tube. Balbiani ring protein 3 could play a
 role as a transport protein that binds to other proteins
 intracellularly and in the gland lumen in order to prevent these
 from forming water-insoluble fibers too early.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Salivary gland.
 CC -!- DOMAIN: Has 82 approximate repeats of Cys-x-Cys-x-cys.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib>)
 or send an email to license@sib-sib.ch.
 CC -----
 DR EMBL; X52263; CAA36506.1; -
 DR PIR; S08167; S08167.
 DR HSSP; P15358; 1SKZ.
 DR InterPro; IPR004153; CXXC_repeat.
 DR Pfam; PF03128; CXXC; 71.
 KW Repeat; Signal; 1 20 POTENTIAL.
 FT SIGNAL 21 1700 BALBIANI RING PROTEIN 3.
 FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.
 SQ SEQUENCE 1700 AA; 186145 MW; 34202829521B0815 CRC64;
 Query Match 8.7%; Score 654.5; DB 1; Length 1700;
 Best Local Similarity 22.3%; Pred. No. 1.1e-26;
 Matches 351; Conservative 30; Mismatches 650; Indels 545; Gaps 89;
 Y 50 GAGTGA-TTTTAAATCTTCTGCGGGA-----GAGGCC 84
 D 94 GCGAGYWCNQDCSKSTPMSAGGCSGSIWEKSCACVCPNADKTAPOVWNRKDTCCC 153
 Y 85 GC-----CTCTC-----CCCGTATCAG----- 102
 D 154 GCPVNNQEPADGCTKPLIWDKCEPLKDKCGKRDWSDSSCSCECKGDKGQSKI 213
 Y 103 -----CGCTTC-----CTCA-----TTCTTTGAA-----TC-CG 125
 D 214 WKNRNCRCICTAPBAGGCSAPLKWDDDKSCACPAKMEKKEKCVESGKIWNPTCEG 273
 Y 126 CGGCTC-----CGCGTCTTCGGCTCAGACGACGCGGAGGAGCCTGTTGC 173
 D 274 CAQLNCPDNKANKETCCQCEKVKKNGGVFKDSCSCVCPGGDKXT----- 323
 Y 174 AATTAGCGGGCTG-----TGACGCGC-----AGGC-----CGCGGGGG 211
 D 324 -----CTAPQVYDVACSCSCPVMNQKPADGCPRPQKWDKEECRC----- 363
 Y 212 CGGCGCGGAGGGGG-----CCATTTTGAATAAGAGGCGTGCCTTC-----CA- 255
 D 364 -----ECPVKHCKNGKVMDETTCQICPRDAPVCTAGKERGSECECKINREPKECAK 419
 Y 256 -----GGCAGG-----CTCTATAAGTGA-----CCGC-C 278
 D 420 PLVNNENTCKVCPADKQMSPGCGSGKSFNKLTCQCECDQSASKGLKRWADTKCEC 479
 Y 279 GCGCGCGAGCG-----TGCGCGCG-TTGCAG-----GTCACCTGTAGCGGACTTCTTTG 325
 D 480 QCPMPPEGCGQTWISDKCKECSPTITQAPQILDNTCEKCPVNNLAQKECKSPRQ 539
 Y 326 GTTTTCTTTCTTTGGG-----GCAC-CTCTGGACTC-----ACTCCC----- 363
 D 540 WTDKSC--LCECSTTPATCEGQTWGEACQCICPSSGDKNCNKKFFDKPSECKCKNKP 597

364 -CAGCATGAGGCG-----TGAGCCCGGTG-----CGCGCTGTACGAGCGGTGT 410
 598 TCTSPQVWDADDECKCKPKDKQKPGCGDGGQWNRVCS-----GCPVPRP 645
 411 GCTGCTGTGGAAGGAGTGTGCG-----CATGCCCGGG-----GCGGAGGAGG 458
 646 DCTNGQIYINWTCAGC-GIDKPCPKQIYINWTKTDCBPCNGMKPVGCGAKTW-LDD 703
 459 GCCC-----GGCAGCTGAGGAGC-CGCTGAGCTTGTCTGG-----AGC-ACATGAAC 502
 704 EQQCDVCPGPKGGTGAQWCDTKCKCKEKMPTGGCENNKWCDDETCDVCPOKNTC 763
 503 CACTGCTACTCCCGCTGCGGGAACCTGGTACCCCGAGTCCGAGAGGC--ACTCAGCTTA 560
 764 TAPKVDARTCSICVNPFPKNSPVILKDTCCCGCQNVKSKAPQKFIENICDA-CPNK 822
 561 GCCAGTGGAAATCTTACAGCGCTCATCGACTA-----CATTTCTCGA-----CCTG-- 607
 823 KQCKAPLWSDPFC-----DCVCPNSASMTCLSPKEMWNVKTCTDCDNPFPKDCPCTQK 877
 608 -----CAGTAGTCTCTGG-----CGAGCCAGCCCCCTGG-----ACCCCTGATGG 648
 878 WMDKCKCGCNAQTDCAGGKKFNDFTCCSCGCPGSKLDCTGNTKWSAETCTCGGQVNRN 937
 649 C-----CCCCAC-----C-----TTCCCATCCAGACAGC-----CGAGC 677
 938 CGNLKNFNDNLQCECKNKQEMANCKSPRTWNYDTCKVCVKNADSDDDCKVPQIWLDDQC 997
 678 TCGCTCCGGAATTTGTC-----ATCTCAAGACAAAGAGCTTTTGCA--CTGACTCGG 732
 998 KGCSPASQMTCPANKRPIEKSCS--ECKSPMSPIPOGKKWNEKCVVECANVKTCE- 1054
 733 CGGTGCTCTGACAC-C-----TCCAG-----AACGACAGTGTCTGG-----CGCCCG 772
 1055 --GPQRWCDNQCKICPQVNTKCSDKOKFTESKCEGCTGTQCKDGRFNSNLECGCLCD 1112
 773 -----TTCTGCTGGACCCCGGA-AC--CTCTC-----CTGCCGAA 808
 1113 DKCKPGKQVFNKNTCCQCKPNQKPGDTGNGKDFPLDCKCKNKPANGCTGQVWNE 1172
 809 GCGCGAC-----GCGAGGATGGGCCCCCACTTCGCCC-----TGCCCACTTGACTT 855
 1173 EKQCECFKPKKQCPGGQDNNHQCQCGCTPAPTCSNNQKYSNVSCGCGNPKPKN 1232
 856 CACCAATCCCTTC-CTGGAGAC-----TAAACT--GGTGTCTAGGAGCGAAGACT 905
 1233 -GCPNQIWCNTCRC-----VCPNMEKPADNCKTKWMDNEMCCQCKPCC--PBGCK 1284
 905 GTGAACCTTGTAGCTGAGAGCCAGAC-----TAGCTC-----TGCC----- 943
 1285 GV-----MKWNANTSCCEPADKAPASCGDKKSWNDDSCQCKSKMPGCGCPNQWNE 1340
 944 -----CACCA--GCTGGGCGAGTCAACC--TG-CTCCACCCACCCCGAGTTCTAAGG 994
 1341 KDCBCKCAGTGNCPAGQTNWNSQTCQSCPATGKTGAQVWCSKACKVCVPAQKKCDSP-K 1399
 995 TCTTTTCAGAGCTGGAGGTGTGAAGAGTGGTGTCTCTCCAACTATGCC----- 1046
 1400 TWENSCS--CQCFKNRPTGCGNAGRWDDAT-----CSEKCAATPKCDSPKVFPD 1450
 1047 AAGCGCGCGGAGAGTGTG-----C-TTCTGCTCTCTTGGAGAAAGTTCTTTCGCC 1100
 1451 TTCGC-KGNPNKLDAGRTWDEQCKMTCDLPAIPCHY--EQVYDSNTCK-----CGC 1503
 1101 TGATTTATG-----AACTTATAATAGATATAGTTTGTACTTTTACAGGAA 1154
 1504 PKEETCKGQFSPKSGCKCILECNKKDPGCGAKKIWCQETCKCECSEPPRMCGPNRY 1563
 1155 GGTGACTTTCTGTAA 1170
 1564 WDKEDACQCLTKVC 1579

[illegible]


```

QY 963 -CCC-----TGCTC-----CCACCCAC-----C-----CCCAAGTTCTAAGGTCTT 998
Db 1192 RCNCPGYTGLHCEADINECREFTCHAAHTRDCLQDPGHPFCICLPFT-----GPRCQT 1247
QY 999 T--TCAGAGCCTGGAGGTGTGAAGAGTGGCT-----GCTCTCCAAACTATGCCAAGCG 1052
Db 1248 ALFPCEQPCHGGQCRPSLGRGGGLTFCHCVQPFNGLRERVARSRELOC----- 1300
QY 1053 GCGGAGAGCTGGTCTTCTGGTCTCT-TGGAGAAAGTTCTGTGCCCTGATTATGAA 1111
Db 1301 -----PVGIFCQ-----QTARGPRCACPPGLSGPSCRVSRASPSGATNTSC-AATPCLHGS 1351
QY 1112 C-----TCTATAATAGATATATAGTTTGTACCTTTTACAGGAAGGTGACTT 1162
Db 1352 CLPVQSVPPFCVCAPGCGPRCETPSAPEVPEPRC-----PRAACQ--AKRGDQNCDR 1405
QY 1163 TCTGTAAACATCGATG 1179
Db 1406 ECN-----SPGCGWDG 1416

```

Search completed: May 7, 2004, 15:08:52
Job time : 35 secs